STOCHASTIC CALCULUS OF VARIATIONS FOR GENERAL LÉVY PROCESSES AND ITS APPLICATIONS TO JUMP-TYPE SDE'S WITH NON-DEGENERATED DRIFT

ALEXEY M.KULIK

Abstract. We consider an SDE in $\mathbb{R}^m$ of the type $dX(t) = a(X(t))dt + dU_t$ with a Lévy process U and study the problem for the distribution of a solution to be regular in various senses. We do not impose any specific conditions on the Lévy measure of the noise, and this is the main difference between our method and the known methods by J.Bismut or J.Picard. The main tool in our approach is the stochastic calculus of variations for a Lévy process, based on the time-stretching transformations of the trajectories.

Three problems are solved in this framework. First, we prove that if the drift coefficient a is non-degenerated in an appropriate sense, then the law of the solution to the Cauchy problem for the initial equation is absolutely continuous, as soon as the Lévy measure of the noise satisfies one of the rather weak intensity conditions, for instance the so-called *wide cone condition*. Secondly, we provide the sufficient conditions for the density of the distribution of the solution to the Cauchy problem to be smooth in the terms of the family of the so-called *order indices* of the Lévy measure of the noise (the drift again is supposed to be non-degenerated). At last, we show that an invariant distribution to the initial equation, if exists, possesses a C^∞-density provided the drift is non-degenerated and the Lévy measure of the noise satisfies the wide cone condition.

Introduction

In this paper, we consider an SDE in $\mathbb{R}^m$ of the type

$$dX(t) = a(X(t))dt + dU_t, \tag{0.1}$$

where $a \in C^1(\mathbb{R}^m, \mathbb{R}^m)$ satisfies the linear growth condition and $U.$ is a Lévy process in $\mathbb{R}^m$. We study the properties of the distribution of both the solution $X(x,\cdot)$ to the Cauchy problem associated with (0.1) and a stationary solution $X^*(\cdot)$ to (0.1), supposing latter to exist. The question under discussion is the following one: do the distributions $P_{x,t}(dy) \equiv P(X(x,t) \in dy), P^*(dy) \equiv P(X^*(t) \in dy)$ of these solutions have densities $p_{x,t}, p^*$ w.r.t. the Lebesgue measure λ^m in $\mathbb{R}^m$? Do these densities possess any additional regularity property, for instance, belong to the class C^∞? This question is a natural analog for the classical hypoellipticity problem for partial differential equations, and it can be reformulated in analytic terms in the following way. Let L be the *Lévy-type* pseudo-differential operator

$$Lf(x) = (\nabla f(x), a(x))_{\mathbb{R}^m} + \int\limits_{\|u\|_{\mathbb{R}^m}>1} \Big[f(x+u) - f(x)\Big]\Pi(du) + \int\limits_{\|u\|_{\mathbb{R}^m}\leq 1} \Big[f(x+u) - f(x) - (\nabla f(x), u)_{\mathbb{R}^m}\Big]\Pi(du)$$

associated with (0.1), where Π is the Lévy measure for U. Then $P_{x,t}(dy)$ is the fundamental solution to the Cauchy problem for the operator $\partial_t - L$ and $P^*(dy)$ is the invariant measure for the operator L.

The hypoellipticity problem for equations of the type (0.1) and the more general equations

$$dX(t) = a(X(t))dt + \int_{\mathbb{R}^m} c(X(t-), u)\tilde{\nu}(dt, du) \tag{0.2}$$

with a compensated Poisson point measure $\tilde{\nu}$ was studied by numerous authors.

2000 *Mathematics Subject Classification.* Primary 60H07; Secondary 60G51.

Key words and phrases. Lévy process, admissible time-stretching transformations, differential grid, stochastic calculus of variations.

First of all, let us mention the analytic approach, see [15] and survey in [16]. This approach uses some version of the parametrix method, and the typical conditions demanded here contain the assumptions on a smoothness and a growth rate of the probability density of the initial process U (roughly speaking, the noise should be close to the one generated by a stable process).

There also exist two groups of probabilistic results inspired by the Malliavin's approach to the hypoellipticity problem in the diffusion (i.e., parabolic) setting. The first group is based on the method, in which a Malliavin-type calculus on the space of the trajectories of Lévy processes is introduced via the transformations of trajectories that change values of their jumps. This approach was proposed by J.Bismut ([3]). In this method the Lévy measure was initially supposed to have some (regular) density w.r.t Lebesgue measure. This is a natural condition sufficient for the transformations, changing values of the jumps, to be admissible. There exists a lot of works in this direction, weakening both the non-degeneracy conditions on coefficients and regularity claims on the Lévy measure, cf. [2], [25],[17]. There also exists a closely related approach based on a version of Yu.A.Davydov's *stratification method*, cf. [6], [7]. One can say that this group of results is based on *a spatial regularity* of the noise, which through either Malliavin-type calculus or stratification method guarantees the regularity of the distribution of the functional under investigation.

Another group of results is based on the approach developed by J.Picard, see [29] and [12],[13]. Here the perturbations of the point measure by adding a point into it are used. Since the single perturbation of such a kind generates not a derivative but a difference operator, one should use an ensemble of such perturbations. Therefore *a frequency regularity* is needed, i.e. limitations on the asymptotic behavior of the Lévy measure at the origin should be imposed.

Our aim is to study the hypoellipticity problem for equation (0.1) in a situation where the conditions imposed on the Lévy measure of the noise are as weak as possible. In particular, the noise is not supposed to possess neither spatial nor frequency regularities.

Three problems are solved in this paper. The first one is concerned with the absolute continuity of the law of the solution to (0.1) with non-degenerated drift. We give a general sufficient condition for the absolute continuity without any restrictions on U. The same problem was solved in [22],[23] for the equation of the type (0.2) with some moment restriction on the jump part, and in [28] for the one-dimensional SDE of the type (0.1).

The second problem is to provide the conditions on the Lévy measure of the noise, which would be sufficient and close to the necessary ones for the smoothness of the density of the law of $X(x,t)$. This problem is unsolved even in the case $a = 0, c(x,u) = u$; for the Lévy process U, the criterion for the distribution of U_t to possess a C^∞-density is not known. We show that if the drift coefficient in equation (0.1) is non-degenerated in an appropriate sense, then for the law of $X(x,t)$ such a criterion can be given in the terms of properly defined *order indices* $\boldsymbol{\rho}_r, r \in \mathbb{N}$ of the Lévy measure Π.

The claim on the drift a to be non-degenerated is least restrictive while the problem of the investigation of the properties of the invariant distribution to (0.1) is considered. Such a claim is very natural since the invariant distribution have to exist, and appears to be sufficient for an invariant distribution to possess the C^∞-density under very mild conditions on the jump noise.

Our approach is motivated by a natural idea that, without any conditions on the Lévy measure of U, there always exist admissible transformations of U changing the moments of jumps, and one can construct some kind of stochastic calculus of variations based on these transformations. This idea is not very new, it was mentioned in the introduction to [29]. We also believe that it was one of the motivations for the construction of an integration-by-parts framework for the pure Poisson process in [5] and [8]. However the detailed version of the calculus of variation, based on the time changing transformations, which would give opportunity to

study m-dimensional SDE's, was not available till the recent papers of the author [22],[23] (the preliminary version of such a calculus was proposed by the author in [19]; the similar approach was proposed in [28] with an application to a one-dimensional SDE of the type (0.1)).

The structure of the paper is the following. In Section 1 we formulate the main results of the paper, in Section 2 we make a detailed discussion of these results and give some sufficient conditions and corollaries. In Section 3 the stochastic calculus for Lévy processes, based on the time-stretching transformations, is introduced. The proofs of the statements about the existence of the density for $P_{x,t}(dy)$, smoothness of this density, and smoothness of the density for $P^*(dy)$ are given in Sections 4, 5 and 6, respectively.

1. Main results

1.1. **Auxiliary definitions and notation.** Before formulating the main results of the paper, let us introduce a notation. Denote, by $S_m = \{v \in \mathbb{R}^m | \|v\|_{\mathbb{R}^m} = 1\}$, a unit sphere in $\mathbb{R}^m$. For $v \in S_m, \varrho \in (0,1)$, denote by $V(v,\varrho) \equiv \{y \in \mathbb{R}^m | |(y,v)_{\mathbb{R}^m} \geq \varrho \|y\|_{\mathbb{R}^m}\}$ the two-sided cone with the axis $\langle v \rangle \equiv \{tv, t \in \mathbb{R}\}$.

Definition 1.1. For $r \in \mathbb{N}$, we define

$$\rho_r(\varrho,\varepsilon) = \left[\varepsilon^r \ln \frac{1}{\varepsilon}\right]^{-1} \cdot \inf_{v \in S_m} \int_{V(v,\varrho)} (|(u,v)_{\mathbb{R}^m}| \wedge \varepsilon)^r \Pi(du), \quad \varepsilon > 0, \quad \boldsymbol{\rho}_r = \lim_{\varrho \to 0+} \liminf_{\varepsilon \to 0+} \rho_r(\varrho,\varepsilon) \in [0,+\infty].$$

We call $\boldsymbol{\rho}_r$ *the upper order index of power* r for the Lévy measure Π. The main role in our considerations plays the index $\boldsymbol{\rho}_2$; we denote this index by $\boldsymbol{\rho}$.

Definition 1.2. Define

$$\vartheta(\varepsilon) = \left[\varepsilon^2 \ln \frac{1}{\varepsilon}\right]^{-1} \cdot \sup_{v \in S_m} \int_{\mathbb{R}^m} (|(u,v)_{\mathbb{R}^m}| \wedge \varepsilon)^2 \Pi(du), \quad \varepsilon > 0, \quad \boldsymbol{\vartheta} = \liminf_{\varepsilon \to 0+} \vartheta(\varepsilon) \in [0,+\infty].$$

We call $\boldsymbol{\vartheta}$ *the lower order index* for the Lévy measure Π. In the one-dimensional case, the definition of the order indices is most simple, since $S_1 = \{-1,+1\}$ and $V(v,\varrho) = \mathbb{R}$ for $v = \pm 1, \varrho \in (0,1)$. In the case $m = 1$, we have

$$\rho_r(\varrho,\varepsilon) = \rho_r(\varepsilon) = \left[\varepsilon^r \ln \frac{1}{\varepsilon}\right]^{-1} \cdot \int_{\mathbb{R}} (|u| \wedge \varepsilon)^r \Pi(du), \quad \vartheta(\varepsilon) = \left[\varepsilon^2 \ln \frac{1}{\varepsilon}\right]^{-1} \cdot \int_{\mathbb{R}} (|u| \wedge \varepsilon)^2 \Pi(du),$$

and $\boldsymbol{\vartheta} = \boldsymbol{\rho}$.

Definition 1.3. The function a belongs to the class $\mathbf{K}_r, r \in \mathbb{N}$, if, for every $\varrho \in (0,1)$, there exists $D = D(a,r,\varrho) > 0$ such that, for every $x \in \mathbb{R}^m$, $v \in S_m$, there exists $w = w(x,v) \in S_m$ with

$$|(a(x+y) - a(x), v)_{\mathbb{R}^m}| \geq D|(y,w)_{\mathbb{R}^m}|^r, \quad y \in V(w,\varrho), \|y\|_{\mathbb{R}^m} \in (-D,D). \tag{1.1}$$

The function a belongs to the class $\mathbf{K}^O_{r,loc}$ ($r \in \mathbb{N}$, O is some open subset of $\mathbb{R}^m$) if, for every $x \in O, \varrho \in (0,1)$, there exists $D = D(a,r,\varrho,x) > 0$ such that, for every $v \in S_m$, there exists $w = w(x,v) \in S_m$ with (1.1) being true. The function a belongs to the classes $\mathbf{K}_\infty$ or $\mathbf{K}^O_{\infty,loc}$, if $\exists r \in \mathbb{N} : a \in \mathbf{K}_r$ or $a \in \mathbf{K}^O_{r,loc}$, respectively.

Example 1.1. a) The function $a \in C^1(\mathbb{R}^m,\mathbb{R}^m)$ belongs to the class $\mathbf{K}^O_{1,loc}$ if, for every $x \in O$, $\det \nabla a(x) \neq 0$.

b) The function $a \in C^1(\mathbb{R}^m,\mathbb{R}^m)$ belongs to the class $\mathbf{K}_1$ if $\sup_{x \in \mathbb{R}^m} \left\|[\nabla a(x)]^{-1}\right\|_{\mathbb{R}^{m \times m}} < +\infty$ and ∇a is uniformly continuous.

c) The function $a \in C^r(\mathbb{R},\mathbb{R})$ belongs to the class $\mathbf{K}_r$ if, for some $R, c > 0$, the inequality $|a'(x)| \geq c$ holds for all x with $|x| > R$, and, for every x, one of the derivatives $a'(x), a''(x), \ldots, a^{(r)}(x)$ differs from 0.

Definition 1.4. The measure Π satisfies *the wide cone condition* if, for every $v \in S_m$, there exists $\varrho = \varrho(v) \in (0,1)$ such that $\Pi(V(v,\varrho)) = +\infty$.

Remarks. 1. For $m=1$, the measure Π satisfies the wide cone condition iff $\Pi(\mathbb{R})=+\infty$.

2. In Definition 1.4, the value of the parameter ρ can be chosen to be independent of v; this follows from the compactness of S_m.

Denote, by $CB^k(\mathbb{R}^m)$, the set of the real-valued functions f on $\mathbb{R}^m$ such that f has k Sobolev derivatives and its k-th derivative is a bounded function on $\mathbb{R}^m, CB^0(\mathbb{R}^m)\equiv L_\infty(\mathbb{R}^m)$. Denote also, by $C_b^\infty(\mathbb{R}^m)$, the set of the real-valued infinitely differentiable functions on $\mathbb{R}^m$ that are bounded together with every their derivative. It is clear that $CB^k(\mathbb{R}^m)\subset C^{k-1}(\mathbb{R}^m)$ and $C_b^\infty(\mathbb{R}^m)=\bigcap_{k=1}^\infty CB^k(\mathbb{R}^m)$.

1.2. **Absolute continuity of the law of $X(x,t)$.** In this subsection, the coefficient a is supposed to belong to $C^1(\mathbb{R}^m,\mathbb{R}^m)$ and to satisfy the linear growth condition.

Theorem 1.1. *Suppose that for a given $x_*\in\mathbb{R}^m$ there exists $\varepsilon_*>0$ such that for arbitrary $v\in\mathbb{R}^m\backslash\{0\}, x\in \bar{B}(x_*,\varepsilon_*)\equiv\{y|\|y-x_*\|\le\varepsilon_*\}$*

$$\Pi\Big(u:(a(x+u)-a(x),v)_{\mathbb{R}^m}\neq 0\Big)=+\infty. \tag{1.2}$$

Then, for every $t>0$,

$$P\circ[X(x_*,t)]^{-1}\ll\lambda^m.$$

This statement is analogous to that of Theorem 3.2 [22], but the moment restriction analogous to condition (1.4) below, that was used in [22], is removed here.

The statement of Theorem 1.1 can be generalized in the following way. Consider the sequence of equations of the type

$$X_n(x,t)=x+\int_0^t a_n(X_n(x,s))\,ds+U_t^n+V_t^n,\quad t\in\mathbb{R}^+, \tag{1.3}$$

where V^n are non-random functions from the Skorokhod's space $\mathbb{D}(\mathbb{R}^+,\mathbb{R}^m)$, and the Lévy processes U^n are given by stochastic integrals

$$U_t^n=U_0+\int_0^t\int_{\|u\|>1}c_n(u)\nu(ds,du)+\int_0^t\int_{\|u\|\le 1}c_n(u)\tilde{\nu}(ds,du),\quad t\in\mathbb{R}^+,n\in\mathbb{N}.$$

Theorem 1.2. *Suppose that the following conditions hold true:*

1) the coefficients $a_n, n\ge 1$ belong to $C^1(\mathbb{R}^m,\mathbb{R}^m)$ and satisfy the uniform linear growth condition;

2) $a_n\to a,\nabla a_n\to\nabla a, n\to+\infty$, uniformly on every compact set;

3) the functions $\|c_n\|$ are dominated by a function $\mathbf{c}$ with $\int_{\mathbb{R}^m}\left[\mathbb{1}_{\|u\|\le 1}\mathbf{c}^2(u)+\mathbb{1}_{\|u\|>1}\mathbf{c}(u)\right]\Pi(du)<+\infty$;

4) $c_n(u)\to u, n\to+\infty$ for Π-almost all $u\in\mathbb{R}^m$;

5) $V^n\to V, n\to+\infty$ in $\mathbb{D}(\mathbb{R}^+,\mathbb{R}^m)$;

6) $x_n\to x_*, t_n\to t>0, n\to+\infty$ and the function V is continuous at the point t.

Suppose also that the function a, the measure Π and the point x_* satisfy the condition of Theorem 1.1.

Then the laws of $X_n(x_n,t_n)$ converge in variation to the law of the solution $X(x_*,t)$ to the equation

$$X(x_*,t)=x_*+\int_0^t a(X(x_*,s))\,ds+U_t+V_t,\quad t\in\mathbb{R}^+.$$

As a corollary, we obtain the following uniform version of Theorem 1.1.

Corollary 1.1. *Suppose that the conditions of Theorem 1.2 hold true. Suppose also that, for every $n\in\mathbb{N}$, the function a_n, the measure $\Pi_n(du)=c_n(u)\Pi(du)$, and the point x_n satisfy the condition of Theorem 1.1, and $t_n>0$. Then the family of the distributions of $X_n(x_n,t_n), n\ge 1$ is uniformly absolutely continuous.*

Let us also give a partial form of the Corollary 1.1, that is important by itself.

Corollary 1.2. *Suppose that the condition of Theorem 1.1 holds true for every $x_* \in \mathbb{R}^m$. Then the map*

$$\mathbb{R}^m \times (0, +\infty) \ni (x,t) \mapsto p_{x,t} \in L_1(\mathbb{R}^m)$$

is continuous, and therefore the process X is strongly Feller.

1.3. **Smoothness of the density $p_{x,t}$.** In this paper, while solving the problem of the smoothness of the density (both of the law of $X(x,t)$ and of the law of $X^*(t)$), we restrict ourselves by the Lévy processes satisfying the following moment condition:

$$\int_{\|u\|_{\mathbb{R}^m}\leq 1} \|u\|_{\mathbb{R}^m} \Pi(du) < +\infty. \tag{1.4}$$

This supposition is crucial for the specific form of the calculus of variations developed below. We believe that this limitation can be removed, and the results given below also holds true for the Lévy processes without any additional moment conditions. But such an expansion should involve some more general version of the calculus of variations, based on a "more singular" integration-by-parts formula. This is a subject for the further investigation.

The coefficient a is supposed to be infinitely differentiable and to have all the derivatives bounded. We also suppose that

$$\int_{\{\|u\|>1\}} \|u\|^p \Pi(du) < +\infty \text{ for every } p < +\infty. \tag{1.5}$$

These conditions imply, in particular, that

$$E \sup_{s\leq t} \|X(x,s) - x\|^p < +\infty, \quad p < +\infty. \tag{1.6}$$

Conditions on the coefficient a and condition (1.5) are technical ones and, unlike condition (1.4), can be replaced by more weak analogs in the formulation of the most of the results given below. In order to make the exposition transparent and reasonably short, we omit these considerations.

The main regularity result is given by the following theorem. Denote $\mathbf{c}(k,m) = \frac{2e}{e-1}(km + m^2 + 2m - 2)$, $k \geq 0, m \in \mathbb{N}$.

Theorem 1.3. *Let $a \in \mathbf{K}_r$ and $\boldsymbol{\rho}_{2r} \in (0, +\infty]$ for some $r \in \mathbb{N}$. Then, for every $x \in \mathbb{R}^m$ and $t \in \mathbb{R}^+$ with $t\frac{\boldsymbol{\rho}_{2r}}{2r} > \mathbf{c}(k,m)$, the density $p_{x,t}$ belongs to the class $CB^k(\mathbb{R}^m)$. In particular, if $a \in \mathbf{K}_r$ and $\boldsymbol{\rho}_{2r} = +\infty$ for some $r \in \mathbb{N}$, then $p_{x,t} \in C_b^\infty(\mathbb{R}^m)$ for every $t \in \mathbb{R}^+$.*

The following theorem shows that the conditions given before are rather precise. Denote, by Θ, the set of (x,t) such that $P(X(x,t) \in dy) = p_{x,t}(y)dy$. We do not claim Θ to coincide with $\mathbb{R}^m \times (0, +\infty)$ and give the properties of $p_{x,t}$ for $(x,t) \in \Theta$.

Theorem 1.4. a. *The density $p_{x,t}$ does not belong to $L_{r,loc}(\mathbb{R}^m)$ for $t\boldsymbol{\vartheta} < m(1 - \frac{1}{r})$, $r > 1$.*

b. *The density $p_{x,t}$ does not belong to $C(\mathbb{R}^m)$ for $t\boldsymbol{\vartheta} < m$.*

If the condition (1.4) fails, then the following analogues of **a**,**b** *hold true:*

a1. *the density $p_{x,t}$ does not belong to $L_r(\mathbb{R}^m)$ for $t\boldsymbol{\vartheta} < m(1 - \frac{1}{r})$;*

b1. *the density $p_{x,t}$ does not belong to $CB^0(\mathbb{R}^m)$ for $t\boldsymbol{\vartheta} < m$.*

1.4. **Smoothness of the invariant distribution.** Like in the previous subsection, the coefficient a is supposed to be infinitely differentiable and to have all the derivatives bounded. The jump noise is claimed to satisfy the moment conditions (1.4), (1.5). Consider the invariant distribution P^* of (0.1) or, equivalently, the distribution of $X^*(t)$, where $X^*(\cdot)$ is a stationary process satisfying (0.1). We suppose the invariant distribution to exist and to have all the moments (we do not claim this distribution to be unique).

Remark. The most simple sufficient condition here is the claim for the drift coefficient a to be "dissipative at the infinity":

$$\exists R \in \mathbb{R}^+, \gamma > 0: \quad (a(x), x)_{\mathbb{R}^m} \le -\gamma \|x\|^2_{\mathbb{R}^m}, \quad \|x\|_{\mathbb{R}^m} \ge R. \tag{1.7}$$

Condition (1.7), together with (1.5), guarantees both that P^* exists and that P^* has all the moments.

Theorem 1.5. *Let Π satisfy the wide cone condition and $a \in \mathbf{K}_\infty$.*

Then $P^*(dy) = p^*(y)dy$ with $p^* \in C_b^\infty(\mathbb{R}^m)$.

2. Sufficient conditions, examples and discussion

In this section, we would like to demonstrate by a detailed discussion the general results formulated in Theorems 1.1 – 1.5.

2.1. **Absolute continuity of the law of $X(x,t)$.** Let us formulate several sufficient conditions for the condition (1.2) to hold true. We are interested in the conditions on the drift a, such that, under minimal assumptions on the jump noise, the solution to (0.1) has the absolutely continuous distribution. Obviously, the necessary assumption here is that $\Pi(\mathbb{R}^m) = +\infty$, because otherwise the distribution of $X(t)$ has an atom.

The first condition is given in the case $m = 1$. Everywhere below x_* is used for the initial value of the solution. Denote $N(a, y) = \{x \in \mathbb{R} | a(x) = y\}$.

Proposition 2.1. *Suppose that $\Pi(\mathbb{R}) = +\infty$ and there exists some $\delta_* > 0$ such that*

$$\forall y \in \mathbb{R} \quad \#\Big[N(a, y) \cap (x_* - \delta_*, x_* + \delta_*)\Big] < +\infty.$$

Then (1.2) holds true, and therefore, for every $t > 0$,

$$P \circ [X(x_*, t)]^{-1} \ll \lambda^1.$$

In [28], in the case $m = 1$ only, the law of $X(t)$ was proved to be absolutely continuous under condition that $a(\cdot)$ is strictly monotonous at some neighborhood of x_*. One can see that this condition is somewhat more restrictive than the one of Proposition 2.1. The proof of Proposition 2.1, as well as the proofs of Propositions 2.2, 2.3 below, is given in the subsection 4.3.

The second sufficient condition is formulated for multidimensional case.

Proposition 2.2. *Let the measure Π satisfy the wide cone condition. Suppose that there exists a neighborhood O of the initial point x_* such that $a \in K^O_{\infty,loc} \equiv \bigcap_r K^O_{r,loc}$.*

Then (1.2) holds true, and therefore, for every $t > 0$,

$$P \circ [X(x_*, t)]^{-1} \ll \lambda^m.$$

One can give some more precise versions of the sufficient condition in the multidimensional case, if the structure of the drift coefficient is specified in more details.

Define a *proper smooth surface* $S \subset \mathbb{R}^m$ as any set of the type $S = \{x | \phi(x) \in L\}$, where L is a proper linear subspace of $\mathbb{R}^m$ and $\phi \in C^1(\mathbb{R}^m, \mathbb{R}^m)$ is such that $\det \nabla \phi(0) \ne 0$ and $\phi^{-1}(\{0\}) = \{0\}$.

Proposition 2.3. *Suppose that one of the following group of conditions holds true:*

a. *$a \in C^1(\mathbb{R}^m, \mathbb{R}^m)$, $\det \nabla a(x_*) \ne 0$ and*

$$\Pi(\mathbb{R}^m \backslash S) = +\infty \quad \textit{for every proper smooth surface } S; \tag{2.1}$$

b. *$a(x) = Ax, A \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)$ is non-degenerate and*

$$\Pi(\mathbb{R}^m \backslash L) = +\infty \quad \textit{for every proper linear subspace } L \subset \mathbb{R}^m. \tag{2.2}$$

Then (1.2) holds true, and therefore, for every $t > 0$,

$$P \circ [X(x_*, t)]^{-1} \ll \lambda^m.$$

Condition (2.1) is less restrictive than the wide cone condition introduced in Definition 1.4. It holds true, for instance, if $\Pi(\mathbb{R}^m \backslash Y) = +\infty$ for every set $Y \subset \mathbb{R}^m$, whose Hausdorff dimension does not exceed $m-1$.

Condition (2.2) is close to the necessary one, this is illustrated by the following simple example. Let (2.2) fail for some L, and let L be invariant for A. Then, for $x_* \in L$ and any $t \geq 0$, $P(X(x_*, t) \in L) > 0$. Therefore, the law of $X(x_*, t)$ is not absolutely continuous.

Condition (2.2) was introduced by M.Yamazato in the paper [41], where the problem of the absolute continuity of the distribution of the Lévy process was studied. This condition obviously is necessary for the law of U_t to possess a density. In [41], some sufficient conditions were also given. Statement 4 of the main theorem in [41] guarantees the absolute continuity of the law of U_t under the following three assumptions:

(a) condition (2.2) is valid;

(b) $\Pi(L) = 0$ for every linear subspace $L \subset \mathbb{R}^m$ with $\dim L \leq m-2$;

(c) the conditional distribution of the radial part of some *generalized polar coordinate* is absolutely continuous.

We would like to note that assumption (c) is some kind of a "spatial regularity" assumption (in the sense we have used in Introduction) and is crucial in the framework of [41]. Without such an assumption, condition (2.2) is not strong enough to guarantee U_t to possess a density, this is illustrated by the following example.

Example 2.1. Let $m = 2, \Pi = \sum_{k\geq 1} \delta_{z_k}$, where $z_k = (\frac{1}{k!}, \frac{1}{(k!)^2}), k \geq 1$. Every point z_k belongs to the parabola $\{z = (x, y) | y = x^2\}$. Since every line intersects this parabola at not more than two points, condition (2.2) together with assumption (b) given before hold true. On the other hand, for any $t > 0$, it is easy to calculate the Fourier transform of the first coordinate U_t^1 of $U_t = (U_t^1, U_t^2)$ and show that

$$\lim_{N\to+\infty} E \exp\{i2\pi N! U_t^1\} = 1.$$

This means that the law of U_t^1 is singular, and consequently the law of U_t is singular too.

Due to Proposition 2.3, (2.2) is the exact condition for the *linear* multidimensional equation (0.1) to possess the same regularization feature with the one given in Introduction. We have seen that the process U_t may satisfy this condition and fail to have an absolutely continuous distribution. However, adding a non-degenerated linear drift, we obtain the solution to (0.1) (i.e., an Ornstein-Uhlenbeck process with the jump noise) with the absolutely continuous distribution. At this time, we cannot answer the question whether (2.2) is strong enough to handle the non-linear case, i.e. whether statement **a** of Proposition 2.3 is valid with (2.1) replaced by (2.2).

2.2. **Smoothness of the density** $p_{x,t}$**.** Theorems 1.3,1.4 allows one to completely describe the regularity properties of the distribution density of the solution to (0.1) in the case $m = 1$. These properties are determined by the value of the order index $\boldsymbol{\rho}$ (remind that for $m = 1$ the upper order index $\boldsymbol{\rho}$ coincides with the lower order index $\boldsymbol{\vartheta}$), the only possible cases here are $\boldsymbol{\rho} = +\infty, \boldsymbol{\rho} = 0, \boldsymbol{\rho} \in (0, +\infty)$.

The case of $\boldsymbol{\rho} = +\infty$ is "diffusion-like", which means that if $a \in \mathbf{K}_1$ then the density $p_{x,t}$ instantly (i.e., for every positive t) becomes infinitely differentiable. The opposite case $\boldsymbol{\rho} = 0$ means that the intensity of the noise is too low to produce the regular density and for every $x \in \mathbb{R}, t \in \mathbb{R}^+, p > 1$ the density $p_{x,t}$, if exists, does not belong to $L_{p,loc}(\mathbb{R})$.

If we compare equation (0.1) with the diffusion equations, an essentially new feature occurs in the intermediate case $\boldsymbol{\rho} \in (0, +\infty)$. On the one hand, if $a \in \mathbf{K}_1$, then we see from Theorem 1.3 that there exists a

sequence $\{\mathbf{a}_k = \frac{2e(k+1)}{\boldsymbol{\rho}(e-1)}, k \geq 0\}$ such that $p_{x,t} \in CB^k(\mathbb{R})$ as soon as $t > \alpha_k$. On the other hand, $p_{x,t} \notin CB^0(\mathbb{R})$ for t small enough. We believe that such a feature was not known before and introduce for it the term *gradual hypoellipticity.*

Thus, if $m = 1$ and $a \in \mathbf{K}_1$, then the only possibilities for the law of $P_{x,t}$ are

- $P_{x,t}$ does not have a density of the class $\bigcup_{p>1} L_{p,loc}$ for any $t > 0$ ($\boldsymbol{\rho} = 0$);
- the density of $P_{x,t}$ becomes C^k-differentiable after some non-trivial period of time ($\boldsymbol{\rho} \in (0, +\infty)$);
- the density of $P_{x,t}$ instantly becomes infinitely differentiable ($\boldsymbol{\rho} = +\infty$).

In some cases the gradual hypoellipticity feature can be described in more details.

Proposition 2.4. *Let $m = 1$ and $\boldsymbol{\rho}_1 < +\infty$. Let Π be one-sided, i.e. $\Pi((-\infty, 0)) \cdot (\Pi(0, +\infty)) = 0$. Then the density $p_{x,t}$ does not belong to $CB^k(\mathbb{R})$ for $t\boldsymbol{\rho}_1 < k + 1$.*

For the proof of Proposition 2.4 see subsection 5.1. If the conditions of this Proposition hold true, $\boldsymbol{\rho} > 0$ and $a \in \mathbf{K}_1$, then the rate of smoothness of the density is increasing gradually: there exist two progressions $\{\mathbf{a}_k = \alpha k + \beta\}$ and $\{\mathbf{b}_k = \gamma k + \delta\}$ ($\alpha, \gamma > 0$) such that $p_{x,t} \notin CB^k$ while $t < \mathbf{b}_k$, but $p_{x,t} \in CB^k(\mathbb{R})$ as soon as $t > \mathbf{a}_k$.

Example 2.2. Let $\Pi = \sum_{n \geq 1} \delta_{\gamma^{-n}}, \gamma > 1$, then $\boldsymbol{\vartheta} = \boldsymbol{\rho} = \boldsymbol{\rho}_1 = \frac{1}{\ln \gamma}$, and the conditions of Proposition 2.4 hold true.

The gradual hypoellipticity feature can also occur in the multidimensional case. If $m > 1$, $\boldsymbol{\vartheta} > 0$, $\boldsymbol{\rho}_{2r} < +\infty$ and $a \in \mathbf{K}_r$ for some $r \in \mathbb{N}$, then, on the one hand, for every $k \in \mathbb{N}$ $p_{x,t} \in CB^k(\mathbb{R}^m)$ while t is large enough, but, on the other hand, for every $p > 1$ $p_{x,t} \notin L_{p,loc}(\mathbb{R}^m)$ while t is small enough.

Let us discuss one more question related to Theorems 1.3, 1.4. In Theorem 1.4, no specific conditions on a are imposed. In particular, we can take $a \equiv 0$ and establish the properties of the distribution of the initial Lévy process U. It is easy to see that any condition involving the order indices cannot provide the distribution of U_t to be singular: if $\Pi(du) = \pi(u)\,du$ and $\Pi(\mathbb{R}) = +\infty$, then the distribution of U_t for every $t > 0$ has a density. On the contrary, due to Theorem 1.4, the condition on $\boldsymbol{\vartheta}$ appears to be the proper type of a necessary condition for the distribution of U_t to have *a regular* density. Take for simplicity $m = 1$ and consider the property

$\boldsymbol{UC_b^\infty}$: for every $t > 0$, the distribution of U_t has the density from the class $C_b^\infty(\mathbb{R})$.

Due to Theorem 1.4, the condition $\boldsymbol{\rho} = +\infty$ is necessary for $\boldsymbol{UC_b^\infty}$ to hold true. On the other hand, it is known (see [14],[36]) that if

$$\lim_{\varepsilon \to 0+} \left[\varepsilon^2 \ln \frac{1}{\varepsilon}\right]^{-1} \int_{\{|u| \leq \varepsilon\}} u^2 \Pi(du) = +\infty, \tag{2.3}$$

then $\boldsymbol{UC_b^\infty}$ holds true. The conditions $\boldsymbol{\rho} = +\infty$ and (2.3) are in fact very similar, since we can rewrite the first one to the form

$$\lim_{\varepsilon \to 0+} \left\{ \left[\varepsilon^2 \ln \frac{1}{\varepsilon}\right]^{-1} \int_{\{|u| \leq \varepsilon\}} u^2 \Pi(du) + \left[\ln \frac{1}{\varepsilon}\right]^{-1} \Pi(|u| > \varepsilon) \right\} = +\infty.$$

However, the following example shows that there exists a non-trivial gap between these two conditions.

Example 2.3. Let $\Pi = \sum_{n \geq 1} n \delta_{\frac{1}{n!}}$. Then, for every $r \in \mathbb{N}$,

$$\boldsymbol{\rho}_r \geq \lim \inf_{\varepsilon \to 0+} \left\{ \left[\ln \frac{1}{\varepsilon}\right]^{-1} \Pi(|u| > \varepsilon) \right\} \geq \lim \inf_{N \to +\infty} \frac{1}{\ln N!} \sum_{n \leq N-1} n \geq \lim \inf_{N \to +\infty} \frac{N(N-1)}{2N \ln N} = +\infty.$$

This means that if the coefficient a belongs to K_r for some $r \in \mathbb{N}$, then the solution of (0.1) possesses the C^∞-density. On the other hand, for any $t > 0$, one has

$$\lim_{N\to+\infty}\Big|E\exp\{i2\pi N!U_t\}\Big| = \lim_{N\to+\infty}\prod_{n>N}\Big|\exp\{tn(e^{\frac{i2\pi N!}{n!}} - 1 - \frac{i2\pi N!}{n!})\}\Big| = 1,$$

thus the law of U_t for every $t > 0$ is singular. This provides the example of the situation where $\boldsymbol{\rho} = +\infty$, but the distribution of U_t for every t is essentially singular in a sense that

$$\lim \sup_{|z|\to+\infty} |\phi_{U_t}(z)| = 1, \tag{2.4}$$

where ϕ_{U_t} is used for the Fourier transform of U_t. Moreover, this provides the following new and interesting feature. We say that the Lévy noise in Example 2.3 possesses some *hidden hypoelipticity* (another new term) in the following sense. The law of U_t for every $t \in \mathbb{R}^+$ is singular due to (2.4). But, for any drift coefficient $a \in \mathbf{K}_\infty$ (that is a rather general non-degeneracy condition on a), the law of the solution to (0.1) possesses the C^∞-density.

2.3. **General overview.** Let us summarize the answers on three questions formulated at the beginning of the Introduction. Let us formulate in a compact form some of the previous results. We omit additional technical conditions in the formulation.

Theorem 2.1. I. *If $a \in \mathbf{K}^{\mathbb{R}^m}_{\infty,loc}$ and Π satisfies the wide cone condition, then, for every $t > 0, x \in \mathbb{R}^m$, $P_{x,t} \ll \lambda^m$.*

II. *If $a \in \mathbf{K}_\infty$ and Π satisfies the wide cone condition, then $P^*(dy) = p^*(y)dy$ with $p^* \in C^\infty_b(\mathbb{R}^m)$.*

III.a. *If $a \in \mathbf{K}_r$ and $\boldsymbol{\rho}_{2r} = +\infty$, then $P_{x,t}(dy) = p_{x,t}(y)dy$ with $p_{x,t} \in C^\infty_b(\mathbb{R}^m)$ for every $t > 0$.*

b. *If $a \in \mathbf{K}_r$ and $\boldsymbol{\rho}_{2r} \in (0, +\infty)$, then $P_{x,t}(dy) = p_{x,t}(y)dy$ with $p_{x,t} \in CB^k(\mathbb{R}^m)$ for every $t > \mathbf{a}_k$.*

c. *If $\boldsymbol{\vartheta} = 0$, then $p_{x,t}$, if exists, does not belong to $L_{p,loc}$ for any $t > 0, p > 1$.*

Let us note that, surprisingly, the sufficient conditions for an invariant distribution to possess *smooth* density (the part II. for Theorem 2.1) look like much more similar to the sufficient conditions for $P_{x,t}$ to possess *some* density (the part I.) than the conditions for $P_{x,t}$ to possess *smooth* density (the part III.).

We would like to finish Section 2 with one more remark. It is known that the property for the distribution of the Lévy process to be absolutely continuous is time-dependent: one can construct a process U_t in such a way that the law of U_t is singular for $t < t_*$ and absolutely continuous for $t > t_*$ for some $t_* > 0$ (see [35],[39] and more recent paper [37]). The results given before show that such a feature is still valid for the solutions of equations of the type (0.1) with non-degenerated drift coefficient, but in a different form. On the one hand, the part I. of Theorem 2.1 shows that the law of $X(x, t)$ is absolutely continuous for every $t > 0$ as soon as $a \in \mathbf{K}^{\mathbb{R}^m}_{\infty,loc}$ and Π satisfies the wide cone condition. Thus the type of the distribution of $X(x, t)$, unlike the one of the distribution of U_t, is not time-dependent. The proper form of such a dependence is the "gradual hypoellipticity" feature. Recall that such a feature occurs when $\boldsymbol{\vartheta} > 0, \boldsymbol{\rho}_{2r} < +\infty$ and $a \in \mathbf{K}_r$ for some $r \in \mathbb{N}$.

Another form of such a dependence is given by parts II., III. of Theorem 2.1, that show that the regularity properties of the distribution density of the stationary solution essentially differ from those of the solution to the Cauchy problem. The stationary solution can be informally considered as the solution to the Cauchy problem with the initial point $-\infty$. Thus one should conclude that while any finite time interval in the case $\boldsymbol{\vartheta} = 0$ is "not long enough" for a non-degenerated drift to generate a smooth density, the infinite time interval is "long enough", provided that a is weakly non-degenerated ($a \in \mathbf{K}_\infty$) and Π satisfies the wide cone condition. These considerations show that the hypoellipticity properties of the solution to (0.1), in general, are essentially time-dependent.

3. Time-stretching transformations and associated stochastic calculus for a Lévy process

3.1. **Basic constructions and definitions.** In this subsection we introduce the stochastic calculus on a space of trajectories of the general Lévy process, that is the basic tool in our approach. This calculus is based on the time-stretching transformations of the jump noise and associated differential structure. Differential constructions of a similar kind have been known for some time, say, the integration-by-parts framework for a pure Poisson process was introduced independently in [5] and [8], some analytic properties of the corresponding differential structure on a configuration space (over $\mathbb{R}^+$ or a Riemannian manifold) were described in a cycle of the papers by N.Privault, cf. [31],[32],[33]. Our construction (introduced initially in [19]) is slightly different and is applicable in the general situation where a spatial variable of the noise is non-trivial. The more detailed exposition, as well as some related notions, such as the joint stochastic derivative and the extended stochastic integral w.r.t. the compensated Poisson point measure, can be found in [21].

Let us introduce the notation. By ν and $\tilde{\nu}$, we denote the point measure and the compensated point measure, involved in the Lévy—Khinchin representation for the process U:

$$U_t = U_0 + \int_0^t \int_{\|u\|>1} u\nu(ds,du) + \int_0^t \int_{\|u\|\le 1} u\tilde{\nu}(ds,du),$$

ν is a Poisson point measure on $\mathbb{R}^+ \times (\mathbb{R}^m\backslash\{0\})$ with the intensity measure $dt\Pi(du)$, $\tilde{\nu}(dt,du) = \nu(dt,du) - dt\Pi(du)$. We use the standard terminology from the theory of Poisson point measures without any additional discussion. The term "(locally finite) configuration" for a realization of the point measure is frequently used. We suppose that the basic probability space $(\Omega, \mathcal{F}, P)$ satisfies condition $\mathcal{F} = \sigma(\nu)$, i.e. every random variable is a functional of ν (or U). This means that in fact one can treat Ω as the configuration space over $\mathbb{R}^+\times(\mathbb{R}^m\backslash\{0\})$ with a respective σ-algebra. Also the notion of the point process $p(\cdot)$ associated with the process U (and the measure ν) is used in the exposition. The domain $\mathcal{D}$ of this process is equal to the (random) set of $t \in \mathbb{R}^+$ such that $U_t \neq U_{t-}$, and $p(t) = U_t - U_{t-}$ for $t \in \mathcal{D}$.

The notation ∇_x for the gradient w.r.t. the space variable x is frequently used. If the function depends only on x, then the subscript x is omitted. If it does not cause misunderstanding, we omit the subscript and write, for instance, $\|x\|$ instead of $\|x\|_{\mathbb{R}^m}$.

Denote $H = L_2(\mathbb{R}^+), H_0 = L_\infty(\mathbb{R}^+) \cap L_2(\mathbb{R}^+), Jh(\cdot) = \int_0^\cdot h(s)\,ds, h \in H$. For a fixed $h \in H_0$, we define the family $\{T_h^t, t \in \mathbb{R}\}$ of transformations of the axis $\mathbb{R}^+$ by putting $T_h^t x, x \in \mathbb{R}^+$ equal to the value at the point $s = t$ of the solution of the Cauchy problem

$$z'_{x,h}(s) = Jh(z_{x,h}(s)), \quad s \in \mathbb{R}, \qquad z_{x,h}(0) = x. \tag{3.1}$$

Since (3.1) is the Cauchy problem for the time-homogeneous ODE, one has that $T_h^{s+t} = T_h^s \circ T_h^t$, and in particular T_h^{-t} is the inverse transformation to T_h^t. Multiplying h by some $a > 0$, we multiply, in fact, the symbol of the equation by a. Now, taking the time change $\tilde{s} = \frac{s}{a}$, we see that $T_h^a = T_{ah}^1, a > 0$, which together with the previous considerations gives that $T_h^t = T_{th}^1, h \in H_0, t \in \mathbb{R}$.

Denote $T_h \equiv T_h^1$, we have just proved that $T_{sh} \circ T_{th} = T_{(s+t)h}$. This means that $\mathcal{T}_h \equiv \{T_{th}, t \in \mathbb{R}\}$ is a one-dimensional group of transformations of the time axis $\mathbb{R}^+$. It follows from the construction that $\frac{d}{dt}|_{t=0} T_{th}x = Jh(x), x \in \mathbb{R}^+$.

Remark. We call T_h the *time stretching transformation* because, for $h \in C(\mathbb{R}^+) \cap H_0$, it can be constructed in a more illustrative way: take the sequence of partitions $\{S^n\}$ of $\mathbb{R}^+$ with $|S_n| \to 0, n \to +\infty$. For every n, we make the following transformation of the axis: while preserving an initial order of the segments, every segment of the partition should be stretched by $e^{h(\theta)}$ times, where θ is some inner point of the segment (if $h(\theta) < 0$ then the segment is in fact contracted). After passing to the limit (the formal proof is omitted here

in order to shorten the exposition) we obtain the transformation T_h. Thus one can say that T_h performs the stretching of every infinitesimal segment dx by $e^{h(x)}$ times.

Denote $\Pi_{fin} = \{\Gamma \in \mathcal{B}(\mathbb{R}^d), \Pi(\Gamma) < +\infty\}$ and define, for $h \in H_0, \Gamma \in \Pi_{fin}$, a transformation T_h^Γ of the random measure ν by

$$[T_h^\Gamma \nu]([0,t]\times\Delta) = \nu([0,T_h t]\times(\Delta\cap\Gamma)) + \nu([0,t]\times(\Delta\backslash\Gamma)), \quad t\in\mathbb{R}^+, \Delta\in\Pi_{fin}.$$

An easy calculation gives that $r_h(t) \equiv \frac{d}{dt}(T_h t) = \int_0^1 h(T_{sh}t)\, ds, t\in\mathbb{R}^+$. We put

$$p_h^\Gamma = \exp\left\{\int_{\mathbb{R}^+} r_h(t)\nu(dt,\Gamma) - \lim_{t\to+\infty}[T_h t - t]\Pi(\Gamma)\right\}.$$

Since $T_h^\Gamma\nu$ is again a random Poisson point measure, its intensity measure can be expressed through $r_h(\cdot), \Pi$ explicitly. Thus the following statement is a corollary of the classical absolute continuity result for Lévy processes, see [38], Chapter 9.

Lemma 3.1. *The transformation T_h^Γ is admissible for the distribution of ν with the density p_h^Γ, i.e., for every $\{t_1,\dots,t_n\}\subset\mathbb{R}^+, \{\Delta_1,\dots,\Delta_n\}\subset\Pi_{fin}$ and the Borel function $\phi:\mathbb{R}^n\to\mathbb{R}$,*

$$E\phi([T_h^\Gamma\nu]([0,t_1]\times\Delta_1),\dots,[T_h^\Gamma\nu]([0,t_n]\times\Delta_n)) = Ep_h^\Gamma\phi(\nu([0,t_1]\times\Delta_1),\dots,\nu([0,t_n]\times\Delta_n)).$$

The statement of the lemma and the fact that $\mathcal{F}$ is generated by ν imply that the transformation T_h^Γ generates the corresponding transformation of the random variables, we denote it also by T_h^Γ.

The image of a configuration of the point measure ν under T_h^Γ can be described in a following way: every point (τ,x) with $x\notin\Gamma$ remains unchanged; for every point $(\tau,x)\in N$ with $x\in\Gamma$, its "moment of the jump" τ is transformed to $T_{-h}\tau$; neither any point of the configuration is eliminated nor any new point is added to the configuration. In a sequel, we suppose that the probability space Ω coincides with the space of locally finite configurations on $\mathbb{R}^+\times\mathbb{R}^d$ and denote, by the same symbol T_h^Γ, the bijective transformation of this space described above.

Let $\mathcal{C}$ be the set of functionals $f\in\cap_p L_p(\Omega,P)$ satisfying the following condition: for every $\Gamma\in\Pi_{fin}$, there exists the random element $\nabla_H^\Gamma f\in\cap_p L_p(\Omega,P,H)$ such that, for every $h\in H_0$,

$$(\nabla_H^\Gamma f, h)_H = \lim_{\varepsilon\to0}\frac{1}{\varepsilon}[T_{\varepsilon h}^\Gamma\circ f - f] \tag{3.2}$$

with convergence in every $L_p, p<+\infty$.

Example 3.1. Let $\Delta\in\Pi_{fin}, f=\tau_n^\Delta\equiv\inf\{t|\nu([0,t]\times\Delta)=n\}$. Then $f\in\mathcal{C}$ and

$$[\nabla_H^\Gamma f](\cdot) = -\mathbf{1}_{[0,\tau_n^\Delta]}(\cdot)\mathbf{1}_{p(\tau_n^\Delta)\in\Gamma}.$$

We denote

$$(\rho^\Gamma,h) = -\int_0^\infty h(t)\,\tilde\nu(dt,\Gamma)$$

and note that $L_p-\lim_{\varepsilon\to0}\frac{p_{\varepsilon h}^\Gamma-1}{\varepsilon} = -(\rho^\Gamma,h), \quad p\in(1,+\infty)$.

Lemma 3.2. *For every $\Gamma\in\Pi_{fin}$, the pair $(\nabla_H^\Gamma,\mathcal{C})$ satisfies the following conditions:*

1) For every $f_1,\dots,f_n\in\mathcal{C}$ and $F\in C_b^1(\mathbb{R}^n)$,

$$F(f_1,\dots,f_n)\in\mathcal{C} \quad \textit{and} \quad \nabla_H F(f_1,\dots,f_n) = \sum_{k=1}^n F_k'(f_1,\dots,f_n)\nabla_H f_k$$

(chain rule).

2) The map $\rho^\Gamma: h\mapsto(\rho^\Gamma,h)$ is a weak random element in H with weak moments of all orders, and

$$E(\nabla_H^\Gamma f,h)_H = -Ef(\rho^\Gamma,h), \quad h\in H, f\in\mathcal{C}$$

(integration-by-parts formula).

3) There exists a countable set $\mathcal{C}_0 \subset \mathcal{C}$ *such that* $\sigma(\mathcal{C}_0) = \mathcal{F}$.

Conditions 1),2) follow from the definition of the class $\mathcal{C}$ and Lemma 3.1; condition 3) holds true due to Example 3.1.

For a given $h \in H, \Gamma \in \Pi_{fin}, p > 1$, consider the map

$$\nabla_h^\Gamma : \mathcal{C} \ni f \mapsto (\nabla_H^\Gamma f, h)_H \in L_p(\Omega, \mathcal{F}, P)$$

as a densely defined unbounded operator. Lemma 3.2 provides that its adjoint operator is well defined on $\mathcal{C} \subset L_q(\Omega, \mathcal{F}, P), \frac{1}{p} + \frac{1}{q} = 1$, by the equality

$$[\nabla_h^\Gamma]^* g = -(\rho^\Gamma, h) g - \nabla_h^\Gamma g.$$

Since $\mathcal{C}$ is dense in $L_q(\Omega, \mathcal{F}, P)$, this means that ∇_h^Γ is closable in the L_p sense.

Definition 3.1. The closure $D_{h,p}^\Gamma$ of ∇_H^Γ in the L_p sense is called the *stochastic derivative* in the direction (h, Γ) of order p. The Γ-stochastic derivative D_p^Γ of order p is defined for $f \in \cap_{h \in H} Dom(D_{h,p}^\Gamma)$ such that there exists $g \in L_p(\Omega, P, H)$ with

$$(g, h)_H = D_{h,p}^\Gamma f, \quad h \in H,$$

by the equality $D_p^\Gamma f = g$. If $p = 2$, then p is omitted in the notation.

Now a differential structure on the initial space of trajectories is constructed, and it is natural to try to develop some calculus which would provide statements of the type "if for a functional f the family $\{D^\Gamma f, \Gamma \in \Pi_{fin}\}$ is non-degenerate in some sense, then the law of f is regular." The stratification method or the Malliavin-type calculus of variations is supposed to be a natural tool here. However, the differential structure developed before has some new specific properties that does not allow us to apply these tools immediately. The most important feature is illustrated by the following example.

Example 3.2. Let $f = \tau_n^\Gamma$, $h, g \in C_b(\mathbb{R}^+) \cap L_2(\mathbb{R}^+)$ be such that $h(t) \int_0^t g(s)\, ds \neq g(t) \int_0^t g(s)\, ds, t > 0$, then

$$D_h^\Gamma D_g^\Gamma f = h(\tau_n^\Gamma) \int_0^{\tau_n^\Gamma} g(s)\, ds \neq g(\tau_n^\Gamma) \int_0^{\tau_n^\Gamma} h(s)\, ds = D_g^\Gamma D_h^\Gamma f$$

almost surely. In particular, this means that the family of transformations $\{T_h^\Gamma, h \in H_0\}$ is not commutative and therefore cannot be considered as an infinite-dimensional additive group of transformations. Roughly speaking, the differential structure described by Γ-stochastic derivative is *non-flat*.

One possible way to overcome this difficulty and to introduce an analog of the stratification method in the framework described before was developed in [22]. There, some transformation (corresponding to the transformation of the Lévy process into the associated point process), that changes the non-flat gradient D^Γ to some linear-type gradient over a space $\mathbb{R}^\infty$, was used. The relation between these two gradients is close to the one between the "damped" and "intrinsic" gradients on the configuration space over the Riemannian manifold (see [33]).

The analysis based on the change of the space and the gradient allows one to apply the stratification method and obtain efficient conditions for the absolute continuity of the distribution of a solution to (0.1) or (0.2). However, this analysis appears to be rather complicated. Below we introduce another approach based on the new notion of a *differential grid*. This approach not only simplifies the way the stratification method can be applied, but also allows us to develop the efficient stochastic calculus of variations and consider the question of the smoothness of the density.

3.2. Differential grids and associated Sobolev classes.

Definition 3.2. A family $\mathcal{G}=\{[a_i,b_i)\subset\mathbb{R}^+,h_i\in H_0,\Gamma_i\in\Pi_{fin},i\in\mathbb{N}\}$ is called *a differential grid* (or simply *a grid*) if

(i) for every $i\neq j$, $\Big([a_i,b_i)\times\Gamma_i\Big)\cap\Big([a_j,b_j)\times\Gamma_j\Big)=\emptyset$;

(ii) for every $i\in\mathbb{N}$, $Jh_i>0$ inside (a_i,b_i) and $Jh_i=0$ outside (a_i,b_i).

Any grid $\mathcal{G}$ generates a partition of some part of the phase space $\mathbb{R}^+\times(\mathbb{R}^m\backslash\{0\})$ of the random measure ν into the cells $\{\mathcal{G}_i=[a_i,b_i)\times\Gamma_j\}$. We call the grid $\mathcal{G}$ *finite*, if $\mathcal{G}_i=\emptyset$ for all indices $i\in\mathbb{N}$ except some finite number of indices.

Denote $T^i_t=T^{\Gamma_i}_{th_i}$. For any $i\in\mathbb{N},t,\tilde t\in\mathbb{R}$, the transformations T^i_t,$T^i_{\tilde t}$ commute because so do the time axis transformations T_{th_i},$T_{\tilde th_i}$. It follows from the construction of the transformations T^Γ_h that, for a given $i\in\mathbb{N},t\in\mathbb{R}$,

$$T^i_t\tau^{\Gamma_i}_n=T_{th_i}\tau^{\Gamma_i}_n\quad\begin{cases}=\tau^{\Gamma_i}_n, & \tau^{\Gamma_i}_n\notin[a_i,b_i)\\ \in[a_i,b_i), & \tau^{\Gamma_i}_n\in[a_i,b_i)\end{cases}\quad\text{for every } n$$

(see Example 3.1 for the notation τ^Γ_n). In other words, T^i_t does not change points of configuration outside the cell $\mathcal{G}_i$ and keeps the points from this cell in it. Therefore, for every $i,\tilde i\in\mathbb{N},t,\tilde t\in\mathbb{R}$, the transformations T^i_t,$T^{\tilde i}_{\tilde t}$ commute, which implies the following proposition. Denote, by $\ell_0\equiv\ell_0(\mathbb{N})$, the set of all sequences $l=\{l_i,i\in\mathbb{N}\}$ such that $\#\{i|l_i\neq0\}<+\infty$.

Proposition 3.1. *For a given grid $\mathcal{G}$ and $l\in\ell_0$, define the transformation*

$$T^{\mathcal{G}}_l=T^1_{l_1}\circ T^2_{l_2}\circ\dots.$$

This definition is correct since the transformation $T^i_{l_i}$ differs from the identical one only for a finite number of indices i. Then $\mathcal{T}^{\mathcal{G}}=\{T^{\mathcal{G}}_l,l\in\ell_0\}$ is the group of admissible transformations of Ω which is additive in the sense that $T^{\mathcal{G}}_{l_1+l_2}=T^{\mathcal{G}}_{l_1}\circ T^{\mathcal{G}}_{l_2},l_{1,2}\in\ell_0$.

It can be said that, by fixing the grid $\mathcal{G}$, we choose, from the whole variety of admissible transformations $\{T^\Gamma_h,h\in H_0,\Gamma\in\Pi_{fin}\}$, the additive family that is more convenient to deal with. Let us introduce the gradients and Sobolev classes associated with such families.

Denote, by ℓ_2, the Hilbert space of the sequences

$$l=\{l_i,i\in\mathbb{N}\}:\quad\|l\|_{\ell_2}\equiv\left[\sum_{i\in\mathbb{N}}l_i^2\right]^{\frac12}<+\infty,\quad(l,\tilde l)_{\ell_2}\equiv\sum_{i\in\mathbb{N}}\sigma_il_i\tilde l_i.$$

Define $\mathbf{l}^i\in\ell_2,i\in\mathbb{N}$ by $\mathbf{l}^i_i=1,\mathbf{l}^i_j=0,i\neq j$.

Definition 3.3. The random element $f\in L_p(\Omega,P,E)$ $(p\in(1,+\infty))$, taking values in a separable Hilbert space E, belongs to the domain of the stochastic derivative $D^{\mathcal{G}}_p$ if

1) for every $i\in\mathbb{N},e\in E$, $(f,e)_E\in Dom(D^{\Gamma_i}_{h_i,p})$;

2) there exists $g\in L_p(\Omega,P,\ell_2\otimes E)$ such that $D^{\Gamma_i}_{h_i,p}(f,e)_E=(g,\mathbf{l}^i)_{\ell_2\otimes E}$ for $e\in E,i\in\mathbb{N}$.

The element g is denoted by $D^{\mathcal{G}}_pf$. If $p=2$, then p is omitted in the notation.

The class of all elements $f\in L_p(\Omega,P,E)$ stochastically differentiable in the sense of Definition 3.3 is denoted by $W^1_p(\mathcal{G},E)$. This class is a Banach space w.r.t. the norm

$$\|f\|^{\mathcal{G},E}_{p,1}\equiv\left\{E\|f\|^p_E+E\Big\|D^{\mathcal{G}}_pf\Big\|^p_{\ell_2\otimes E}\right\}^{\frac1p}$$

since the operator $D^{\mathcal{G},E}_p$ is closed in L_p.

Similarly, define the Sobolev class $W^d_p(\mathcal{G},E)$ for $d\geq 1, p>1$ as the domain of the operator $[D^{\mathcal{G}}_p]^k f$, it is a Banach space w.r.t. the norm

$$\|f\|^{\mathcal{G},E}_{p,d}\equiv\left\{E\|f\|^p_E+E\sum_{k=1}^d\Big\|[D^{\mathcal{G}}_p]^k f\Big\|^p_{[\ell_2]^{\otimes k}\otimes E}\right\}^{\frac{1}{p}}.$$

At last, define $I^{\mathcal{G}}_p$ as the adjoint operator to $D^{\mathcal{G}}_p$. This operator is called the *stochastic integral*, which is natural, in particular, due to the following example (see also [21], Theorems 1.1 and 1.2).

Example 3.3. It follows from Lemma 3.2 that a non-random element $\mathbf{l}^i\in\ell_2$ belongs to the domain of every $I^{\mathcal{G}}_p$, and

$$I^{\mathcal{G}}_p(\mathbf{l}^i)=-\rho^{\Gamma_i}_{h_i}=\int_{(a_i,b_i)\times\Gamma_i}h_i(s)\tilde\nu(ds,du).$$

The following properties of $D^{\mathcal{G}}_p$,$I^{\mathcal{G}}_p$ are due to the chain rule (Lemma 3.2, statement 1). The proof is analogous to the proof of the same properties of the stochastic derivative and integral w.r.t. the Wiener process and is omitted.

Lemma 3.3. *1) Let $f_j\in W^1_p(\mathcal{G},E_j), j=1,\dots,n$, $F:E_1\times\dots\times E_n\to E$ be Frechet differentiable, continuous, and bounded together with its derivative. Then $F(f_1,\dots,f_n)\in W^1_p(\mathcal{G},E)$ and*

$$D^{\mathcal{G}}_pF(f_1,\dots,f_n)=\sum_{j=1}^nF'_j(f_1,\dots,f_n)\cdot D^{\mathcal{G}}_pf_j.$$

2) Let $g\in Dom(I^{\mathcal{G}}_{p_1}), f\in W_{p_2}(\mathcal{G},\mathbb{R})$, $p_1>p_2$. Then $fg\in Dom(I^{\mathcal{G}}_p)$, where $p=\frac{p_1q_2}{p_1q_2-p_1-q_2}, q_2=\frac{p_2}{p_2-1}$, and

$$I^{\mathcal{G}}_p(fg)=f\cdot I^{\mathcal{G}}_{p_2}(g)-(D^{\mathcal{G}}_{p_1}f,g)_{\ell_2}.$$

3.3. **Existence of the density via the stratification method.** In this subsection, we give two sufficient conditions for the existence of the density for a functional on $(\Omega,\mathcal{F},P)$. The first condition is formulated in terms of the Sobolev-type stochastic derivative introduced in the previous subsection.

Theorem 3.1. *Consider the $\mathbb{R}^m$-valued random vector $f=(f_1,\dots,f_m)$ which belongs for some grid $\mathcal{G}$ to $W^1_2(\mathcal{G},\mathbb{R}^m)$. Denote, by $\Sigma^{f,\mathcal{G}}=(\Sigma^{f,\mathcal{G}}_{k,r})^m_{k,r=1}$, the Malliavin matrix for f,*

$$\Sigma^{f,\mathcal{G}}_{k,r}\equiv(D^{\mathcal{G}}f_k,D^{\mathcal{G}}f_r)_{\ell_2},\quad k,r=1,\dots,m,$$

and put $\mathcal{N}(f,\mathcal{G})=\{\omega|\Sigma^{f,\mathcal{G}}(\omega)$ is non-degenerate$\}$. Then

$$P\Big|_{\mathcal{N}(f,\mathcal{G})}\circ f^{-1}\ll\lambda^m.$$

The proof is made in the framework of the stratification method (see [7], Chapter 2 for the basic constructions of this method) and contains several standard steps. First, let us choose a countable set $\ell_*\subset\ell_0$ dense in ℓ_2. For any $\bar l=(l^1,\dots,l^m)\in[\ell_*]^m$, we denote

$$\mathcal{N}(f,\bar l)=\{\omega|\text{ the matrix }\left((D^{\mathcal{G}}f_k,l^r)_{\ell_2}\right)^m_{k,r=1}\text{ is non-degenerate}\}.$$

Then $\mathcal{N}(f,\mathcal{G})=\cup_{\bar l\in[\ell_*]^m}\mathcal{N}(f,\bar l)$ and thus, in order to prove the statement of the theorem, it is sufficient to prove that, for every fixed $\bar l\in[\ell_0]^m$,

$$P\Big|_{\mathcal{N}(f,\bar l)}\circ f^{-1}\ll\lambda^m. \tag{3.3}$$

The set $\bar l$ generates the commutative group of admissible transformations of $(\Omega,\mathcal{F},P)$, indexed by $\mathbb{R}^m$:

$$T_t\equiv T^{\mathcal{G}}_{t_1l^1}\circ\dots\circ T^{\mathcal{G}}_{t_ml^m},\quad t=(t_1,\dots,t_m).$$

In order to prove (3.3), we proceed in the following way. Consider the stratification of $(\Omega, \mathcal{F}, P)$ on the orbits of the group $\{T_t, t \in \mathbb{R}^m\}$, which can be considered in our case after a proper parametrization as $\mathbb{R}^m$ or some proper linear subspaces of $\mathbb{R}^m$. The group $\{T_t\}$ generates a measurable parametrization of $(\Omega, \mathcal{F}, P)$ (the detailed exposition will be given further), and thus P can be decomposed into a regular family of conditional distributions such that every conditional distribution is supported by some orbit. Denote, by $\rho_{\bar{l}} \equiv (\sum_{i\in\mathbb{N}} l_i^1 \rho_{h_i}^{\Gamma^i}, \dots, \sum_{i\in\mathbb{N}} l_i^m \rho_{h_i}^{\Gamma^i})$, the logarithmic derivative of P w.r.t. $\{T_t\}$. Then, for almost all orbits γ, the conditional distribution P_γ, supported by the orbit γ, possess the logarithmic derivative $\rho_{\bar{l},\gamma}$, that is equal to the restriction of $\rho_{\bar{l}}$ on the orbit γ. Since $\rho_{\bar{l}}$ has an exponential moment, $\rho_{\bar{l},\gamma}$ has such a moment too for almost all γ. This implies (see [4], Proposition 4.3.1) that, for almost all γ, P_γ possesses a positive continuous density.

On the almost every orbit γ, the function f_γ is equal to the restriction of f on γ and belongs to the Sobolev class $\cap_p W_p^1(P_\gamma)$. This fact is more or less standard and we do not give the proof here. In a linear framework, this subject was discussed in details in [20]. The non-linear case of a commutative admissible group $\{T_t\}$ is quite analogous. We refer the interested reader to [20] and references therein.

Taking into account this analytic background, we can apply the change-of-variables formula on the almost every orbit γ and obtain the absolute continuity of the image of the measure P_γ under the map f_γ. After all, (3.3) is obtained by the Fubini theorem. We omit this part of the exposition, referring the reader to [7], Chapter 2, or [30].

Now let us verify that our specific group $\{T_t\}$ generates a measurable parametrization of $(\Omega, \mathcal{F}, P)$, i.e. there exists a measurable map $\Phi : \Omega \to \mathbb{R}^m \times \tilde{\Omega}$ such that $\tilde{\Omega}$ is a Borel measurable space and the image of every orbit of the group $\{T_t\}$ under Φ has the form $L \times \{\tilde{\omega}\}$, where L is a linear subspace of $\mathbb{R}^m$. This condition was supposed to hold true under the considerations made before.

In order to shorten the notation, we restrict ourselves to the case where

$$l_i^r = \begin{cases} 1, & i = r \\ 0, & \text{otherwise} \end{cases}, \quad r = 1, \dots, m,$$

the general case is quite analogous. For $i = 1, \dots, m$, we denote $\mathcal{D}_i = \{\tau \in \mathcal{D} \cap (a_i, b_i) | p(\tau) \in \Gamma_i\}$, $c_i = \frac{b_i - a_i}{2}$. Let $\omega \in \Omega$ be fixed. We recall that ω is interpreted as a (locally finite) configuration. Set $I(\omega) = \{i | \mathcal{D}_i(\omega) \neq \emptyset\}$ and, for $i \in I(\omega)$, we define $\tau_i(\omega) = \inf \mathcal{D}_i(\omega)$. Note that, due to condition (ii) of Definition 3.2 for every $i = 1, \dots, m$ and $x \in (a_i, b_i)$, the transformation

$$\mathbb{R} \ni z \mapsto T_{zh_i} x$$

is strictly monotonous and its image is equal to (a_i, b_i). Therefore, for every $i \in I(\omega)$ there exists the unique $z_i(\omega) \in \mathbb{R}$ such that $T_{z_i}^i \tau_i = c_i$. Denote $z(\omega) = (z_1(\omega), \dots, z_m(\omega)) \in \mathbb{R}^m$, where $z_i(\omega) = 0$ for $i \notin I(\omega)$. Denote by $\tilde{\Omega}$ the set of all configurations satisfying the following additional condition: for every cell $\mathcal{G}_i, i = 1, \dots, m$, either the configuration is empty in this cell, or the moment of the first jump in this cell is equal to c_i. Now put, for every $\omega \in \Omega$, $\varpi(\omega) = [T_{z(\omega)}\omega] \in \tilde{\Omega}$. Then the map

$$\Phi : \omega \mapsto (z(\omega), \varpi(\omega))$$

provides the needed parametrization. The theorem is proved.

Another version of the previous result can be given in the terms of the *almost sure* stochastic derivative. Although we will not use the framework of almost sure stochastic derivatives while studying equation (0.1), it can be very useful while studying the distributions of some other classes of functionals. Thus we formulate briefly the main points of this framework.

Definition 3.4. For a given grid $\mathcal{G}$, the functional f is called to be almost surely (a.s.) differentiable w.r.t. $\mathcal{G}$, if there exists a random element $\tilde{D}^{\mathcal{G}}f$ with values in ℓ_2 such that, for every $l\in\ell_0$,

$$\frac{1}{t}\Big[f\circ T_l^{\mathcal{G}}-f\Big]\to(\tilde{D}^{\mathcal{G}}f,l)_{\ell_2},\quad t\to0\quad\text{almost surely.}$$

The element $D^{\mathcal{G}}f$ is called the almost sure (a.s.) derivative of f w.r.t. $\mathcal{G}$.

Theorem 3.2. *Consider the random vector $f=(f_1,\dots,f_m)$ such that, for some grid $\mathcal{G}$, every functional $f_r, r=1,\dots,m$ is a.s. differentiable w.r.t. $\mathcal{G}$. Denote $\tilde{\Sigma}^{f,\mathcal{G}}=(\tilde{\Sigma}^{f,\mathcal{G}}_{k,r})_{k,r=1}^m$,*

$$\tilde{\Sigma}^{f,\mathcal{G}}_{k,r}\equiv(\tilde{D}^{\mathcal{G}}f_k,\tilde{D}^{\mathcal{G}}f_r)_{\ell_2},\quad k,r=1,\dots,m,$$

and put $\tilde{\mathcal{N}}(f,\mathcal{G})=\{\omega|\tilde{\Sigma}^{f,\mathcal{G}}(\omega)$ is non-degenerate$\}$. Then

$$P\Big|_{\tilde{\mathcal{N}}(f,\mathcal{G})}\circ f^{-1}\ll\lambda^m.$$

Proof. Due to the arguments given in the proof of the previous theorem, it is sufficient to prove the same statement in a finite-dimensional case, i.e. when Ω is $\mathbb{R}^m$ and $\mathcal{T}^{\mathcal{G}}$ is the canonical group of linear shifts in $\mathbb{R}^m$. In this situation the needed statement holds true due to the standard change-of-variables formula and the following lemma.

Lemma 3.4. *Let, for some $m,n\in\mathbb{N}$, the function $F:\mathbb{R}^m\to\mathbb{R}^n, G:\mathbb{R}^m\to\mathbb{R}^m\times\mathbb{R}^n$ be such that, for every $a\in\mathbb{R}^m$ for λ^m-almost all $x\in\mathbb{R}^m$,*

$$\frac{1}{t}\|F(x+ta)-F(x)-t(G(x),a)_{\mathbb{R}^m}\|_{\mathbb{R}^n}\to0,\quad t\to0.$$

Then, for every $\varepsilon>0$, there exists $F_\varepsilon\in C^1(\mathbb{R}^m,\mathbb{R}^n)$ such that

$$\lambda^m(\{x|F(x)\neq F_\varepsilon(x)\}\cup\{x|G(x)\neq\nabla F_\varepsilon(x)\})<\varepsilon.$$

This result is a straightforward consequence of the Lebesgue theorem about the points of density for a measurable set and the following statements.

Proposition 3.2. *I. ([9], Theorem 3.1.4). Let the function $f:\mathbb{R}^m\to\mathbb{R}^n$ be approximatively differentiable at every point of a set $A\subset\mathbb{R}^m$ along all the vectors from the basis. Then, for λ^m-almost all points $a\in A$, the function f has the approximative derivative at a.*

II. ([9], Theorem 3.1.16). Let $A\subset\mathbb{R}^m, f:A\to\mathbb{R}^n$ and

$$ap\limsup_{x\to a}\frac{\|f(x)-f(a)\|_{\mathbb{R}^n}}{\|x-a\|_{\mathbb{R}^m}}<+\infty\tag{3.4}$$

for λ^m-almost all $a\in A$. Then, for every $\varepsilon>0$, there exists $g\in C^1(\mathbb{R}^m,\mathbb{R}^n)$ such that

$$\lambda^m(\{x|f(x)\neq g(x)\}<\varepsilon.$$

We are not going to discuss definitions of the approximative limit and derivative here, referring the reader to [9]. Let us only mention that the usual differentiability along some direction implies the approximative differentiability along this direction, and if the approximative derivative exists, then (3.4) holds true. Theorem 3.2 is proved.

The following theorem gives the convergence in variation of the distribution of random vectors in terms of their derivatives, and will be used in the proof of Theorem 1.2.

Theorem 3.3. *For some given grid* $\mathcal{G}$ *and* $p > m$, *consider the sequence of* $\mathbb{R}^m$*-valued random vectors* $\{f^n\} \subset W^1_p(\mathcal{G}, \mathbb{R}^m)$ *such that*

$$f_n \to f \ \text{in} \ W^{\mathcal{G},\mathbb{R}^m}_{p,1}, \quad n \to +\infty.$$

Then, for every $A \subset \mathcal{N}(f, \mathcal{G})$,

$$P\Big|_A \circ f_n^{-1} \to P\Big|_A \circ f^{-1}, \quad n \to +\infty$$

in variation.

The statement of the theorem follows, via the stratification arguments analogous to those given in the proof of Theorem 3.1, from the finite-dimensional criterion for the convergence in variation of the sequence of induced measures, given in [1] (see [1], Theorem 2.1 and Corollary 2.7).

Let us mention that the analog of Theorem 3.3 can be also given in the terms of the almost sure derivatives, but an additional uniform condition on the sequence $\{f_n\}$ should be imposed in this case. We do not discuss this subject here, referring the interested reader to [24].

4. Absolute continuity of the distribution of a solution to an SDE with jumps

4.1. **Differential properties of the solution to an SDE with jumps.** We are going to apply the general results about the existence of the density obtained in the previous section to the specific class of functionals: solutions to SDE's with jumps. The first step, that is necessary here, is to verify whether such solutions are either stochastically or a.s. differentiable. In this subsection, we give the answer to this question.

Consider the Cauchy problem for equation (0.1) of the type

$$X(x,t) = x + \int_0^t a(X(x,s))\, ds + U_t - U_0, \quad t \in \mathbb{R}^+. \tag{4.1}$$

We suppose that a belongs to $C^1(\mathbb{R}^m, \mathbb{R}^m)$. We also impose the linear growth condition on a:

$$\exists K : \|a(x)\|^2 \le K(1 + \|x\|^2).$$

These conditions provide that equation (4.1) has the unique strong solution. Moreover, these solutions considered for different x, t form a *stochastic flow of diffeomorphisms*.

Denote $\Delta(x,u) = a(x+u) - a(x), x \in \mathbb{R}^m, u \in \mathbb{R}^d$.

Theorem 4.1. ***I.*** *For every* $x \in \mathbb{R}^m, t \in \mathbb{R}^+, \Gamma \in \Pi_{fin}, h \in H_0$, *every component of the vector* $X(x,t)$ *is a.s. differentiable w.r.t.* $\{T^\Gamma_{rh}, r \in \mathbb{R}\}$, *i.e. there exist a.s. limits*

$$Y_k(x,t) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[T^\Gamma_{\varepsilon h} X_k(x,t) - X_k(x,t)], \quad k = 1, \dots, m.$$

The process $Y(x, \cdot)$ *satisfies the equation*

$$Y(x,t) = \int_0^t \int_\Gamma \Delta(X(x,s-),u) Jh(s)\, \nu(ds,du) + \int_0^t [\nabla a](X(x,s))Y(x,s)\, ds, \quad t \ge 0. \tag{4.2}$$

II. *The solution* $X(x,t)$ *is stochastically differentiable with the derivative given by (4.2).*

Remark. In a sequel, we use only statement II. Statement I provides here the main part of the proof and is emphasized only for the convenience of the reader.

Remark. The statement close to statement I was proved in [22]. The statement close to statement II was proved in [28] for $m = 1$. We cannot use straightforwardly the result from [28] since the proof there contains some specifically one-dimensional features such as an exponential formula for the derivative of the flow corresponding to the solution of the ODE (Lemma 1 [28]).

Proof of statement I. It is sufficient to consider only the case where $a, \nabla a$ are bounded. The general case follows from this one due to the standard localization arguments.

Denote $\mathcal{D}^\Gamma = \{\tau \in \mathcal{D} : p(\tau) \in \Gamma\}$, $\Omega_k = \{\mathcal{D} \cap \{0,t\} = \emptyset, \#(\mathcal{D}^\Gamma \cap (0,t)) = k, \}, k \geq 0$. Since $\Gamma \in \Pi_{fin}$, $\Omega = \cup_k \Omega_k$ almost surely and it is enough to verify that the needed statement holds true a.s. on every Ω_k. The case $k = 0$ is trivial.

Denote $\nu^*(t, A) = \nu(t, A \backslash \Gamma), U_t^* = \int_0^t \int_{\mathbb{R}^d \backslash \Gamma} u\tilde{\nu}(ds, du)$. For a given $t > 0, \tau \in (0,t), p \in \mathbb{R}^d, x \in \mathbb{R}^m$, consider the process $X_\cdot^\tau$ on $[0,t]$ such that

$$X_t^\tau = \begin{cases} x + \int_0^t a(X_s^\tau)\, ds + U_t^*, & t < \tau \\ x + \int_0^t b(X_s^\tau)\, ds + p + U_t^*, & t \geq \tau \end{cases}.$$

Note that the point process $\{p(T), T \in \mathcal{D}^\Gamma\}$ is independent of ν^*, and the distribution of the variable $\tau_1^\Gamma \equiv \min \mathcal{D}^\Gamma$, while this variable is restricted to Ω_1, is absolutely continuous. Then statement I on Ω_1 follows immediately from Example 3.1 and the following lemma.

Lemma 4.1. *With probability 1 for λ^1-almost all $\tau \in (0,t)$,*

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} X_t^{\tau+\varepsilon} = -\Delta(X_{\tau-}^\tau, p)\mathcal{E}_t^*,$$

where $\mathcal{E}^*$ is the stochastic exponent defined by the equation

$$\mathcal{E}_r^* = I_{\mathbb{R}^m} + \int_\tau^r \nabla b(X^\tau(s))\mathcal{E}_s^*\, ds, \quad r \geq \tau.$$

Proof. X_t^τ is the value at the point t of the solution to the equation

$$d\tilde{X}_t = a(\tilde{X}_t)\, dt + dU_t^*, \tag{4.3}$$

with the starting point τ and the initial value

$$X_\tau^\tau = X_{\tau-}^\tau + p.$$

Suppose that $\varepsilon < 0$. Then $X_s^\tau = X_s^{\tau+\varepsilon}, s < \tau + \varepsilon$. Thus $X_t^{\tau+\varepsilon}$ is also the value of the solution to the same equation with the same starting and terminal points and with the initial value being equal to

$$X_{(\tau+\varepsilon)-}^\tau + p + \int_{\tau+\varepsilon}^\tau a(X_s^{\tau+\varepsilon})\, ds + [U_\tau^* - U_{\tau+\varepsilon}^*].$$

Thus the difference $\Phi(\tau, \varepsilon)$ between the initial values for $X_t^{\tau+\varepsilon}, X_t^\tau$ is equal to $\int_{\tau+\varepsilon}^\tau [a(X_s^{\tau+\varepsilon}) - a(X_s^\tau)]\, ds$.

The process $\{U_t^*\}$ has càdlàg trajectories, and therefore almost surely the set of discontinuities for its trajectories is at most countable. Therefore almost surely there exists the set $\mathbb{T} = \mathbb{T}(\omega) \subset \mathbb{R}^+$ of the full Lebesgue measure such that

$$\delta(t, \gamma) \equiv \sup_{|s-t| \leq \gamma} [\|U_s^* - U_t^*\|] \to 0, \quad \gamma \to 0, \quad t \in \mathbb{T}.$$

Then, for $s \in (\tau + \varepsilon, \varepsilon)$,

$$\|X_s^\tau - X_{\tau-}^\tau\| + \|X_s^{\tau+\varepsilon} - X_{\tau-}^\tau - p\| \leq C_\bullet \{|\varepsilon| + \delta(\tau, |\varepsilon|)\}.$$

Here and below, we denote, by $C_\bullet$, any constant such that it can be calculated explicitly, but its exact form is not needed in a further exposition. Thus, for $\tau \in \mathbb{T}$,

$$\|\Phi(\tau, \varepsilon) + \varepsilon[a(X_{\tau-}^\tau + p) - a(X_{\tau-}^\tau)]\| \leq C_\bullet |\varepsilon|\{|\varepsilon| + \delta(\tau, |\varepsilon|)\},$$

which implies the needed statement.

The case $\varepsilon > 0$ is analogous, let us discuss it briefly. Again, take $\tau \in \mathbb{T}$ and represent X^τ_t as the solution to (4.3) with the initial value $X^\tau_{\tau-} + p$. $X^{\tau+\varepsilon}_t$ is also the solution to (4.3) but with the other starting point $\tau + \varepsilon$. The estimates analogous to ones made before show that, up to the $o(|\varepsilon|)$ terms,

$$X^{\tau+\varepsilon}_{\tau+\varepsilon} - X^\tau_{\tau+\varepsilon} = \varepsilon\Big\{-a(X^\tau_{\tau-} + p) + a(X^\tau_{\tau-})\Big\},$$

which implies the statement of the lemma. The lemma is proved.

Now let $k > 1$ be fixed. Consider the countable family $\mathcal{Q}_k$ of the partitions $Q = \{0 = q_0 < q_1 \dots < q_k = t\}$ with $q_1, \dots, q_{k-1} \in \mathbb{Q}$ and denote

$$\Omega_Q = \{\mathcal{D} \cap \{q_i, i = 0, k\} = \emptyset, \mathcal{D}^\Gamma \cap (q_{i-1}, q_i) = 1, i = 1, \dots, k\}, \quad Q \in \mathcal{Q}_k.$$

We have $\Omega_k = \cup_{Q\in\mathcal{Q}_k}\Omega_Q$. Therefore it is enough to verify the statement of Theorem 4.1 on Ω_Q for a given Q. The distributions of the variables $\tau^\Gamma_j, j = 1, \dots, k$ (see Example 3.1 for the notation τ^Γ_j), while these variables are restricted to Ω_k, are absolutely continuous. Then statement I on Ω_Q follows immediately from Example 3.1, the standard theorem about differentiation of the solution to equation (4.1) w.r.t. the initial value, and the statements analogous of one of Lemma 4.1 and written on the intervals $[0, q_1], [q_1, q_2], \dots, [q_{k-1}, t]$. Statement I is proved.

Proof of statement II. Again, suppose first that $a, \nabla a$ are bounded. In the framework of Lemma 4.1, one has the estimate

$$\|X^{\tau+\varepsilon}_t - X^\tau_t\| \le C_\bullet\, |\varepsilon| \tag{4.4}$$

valid point-wise. Indeed, both $X^{\tau+\varepsilon}_t$ and X^τ_t are the solutions to (4.3) with the same initial point (τ for $\varepsilon < 0$ and $\tau + \varepsilon$ for $\varepsilon > 0$) and different initial values. The difference between the initial values are estimated by

$$\left\|\int_{\tau+\varepsilon}^{\tau} [a(X^{\tau+\varepsilon}_s) - a(X^\tau_s)]\, ds\right\| \le -2\|a\|_\infty \varepsilon$$

for $\varepsilon < 0$ and by

$$\left\|\int_{\tau}^{\tau+\varepsilon} [a(X^{\tau+\varepsilon}_s) - a(X^\tau_s)]\, ds\right\| \le 2\|a\|_\infty \varepsilon$$

for $\varepsilon > 0$. Thus, inequality (4.4) follows from the Gronwall lemma. Using the described before technique, involving partitions $Q \in \mathcal{Q}_k$, and applying the Gronwall lemma once again, we obtain that almost surely on the set Ω_k

$$\|T^\Gamma_{\varepsilon h} X(x, t) - X(x, t)\| \le kC_\bullet\, |\varepsilon|.$$

This means that the family $\{\frac{1}{\varepsilon}[T^\Gamma_{\varepsilon h} X(x, t) - X(x, t)]\}$ we already have proved to converge to the solution to (4.2) almost surely as $\varepsilon \to 0$ is dominated by the variable

$$C_\bullet \cdot \nu(t, \Gamma) \in \cap_p L_p(\Omega, \mathcal{F}, P).$$

Therefore the convergence holds true also in the L_p sense for any p, and $X(x, t)$ is stochastically differentiable with the derivative given by (4.2).

The last thing we need to do is to remove the claim on a to be bounded. Consider a sequence $\{a_n\} \subset C^1_b(\mathbb{R}^m, \mathbb{R}^m)$ such that $a_n(x) = a(x)$ for $\|x\| \le n$. We have just proved that the solution $X_n(x, t)$ to an equation of the type (4.1) with a replaced by a_n is stochastically differentiable and its derivative $Y_n(x, t)$ is given by an equation of the type (4.2) with a replaced by a_n. The sequence $\{a_n\}$ can be chosen in such a way that it satisfies the linear growth condition uniformly w.r.t. n. Under such a choice,

$$X_n(x, t) \to X(x, t), \quad Y_n(x, t) \to Y(x, t), \quad n \to +\infty$$

in every $L_p(\Omega, P, \mathbb{R}^m)$. Since the stochastic derivative is a closed operator, this implies the needed statement for $X(x, t)$. The theorem is proved.

4.2. **The proofs of Theorems 1.1, 1.2.** The proof of Theorem 1.1 is an essentially simplified version of the proof of the analogous statement in [22]. It is based on the other version of the absolute continuity result, with the conditions formulated in the terms of the point process $\{p(\tau), \tau \in \mathcal{D}\}$. Below the initial value x_* is fixed, and we omit it in the notation writing $X(s) \equiv X(x_*, s)$.

Denote, by $\{\mathcal{E}_r\}$, the stochastic exponent, i.e. the $m \times m$-matrix-valued process satisfying the equation

$$\mathcal{E}_r = I_{\mathbb{R}^m} + \int_0^r \nabla a(X(s))\mathcal{E}_s\, ds, \quad r \in \mathbb{R}^+.$$

This process has continuous trajectories. The matrix $\mathcal{E}_r$ is a.s. invertible for every r, and, moreover, almost surely

$$\sup_{r \le t} \|\mathcal{E}_r^{-1}\|_{\mathbb{R}^m \times \mathbb{R}^m} < +\infty.$$

We do not discuss this fact in details, since the technique is quite standard here (see, for instance [34], Chapter 5, §10).

Lemma 4.2. *Denote by S_t a linear span of the set of vectors $\{\mathcal{E}_\tau^{-1} \cdot \Delta(X(\tau-), p(\tau)), \tau \in \mathcal{D} \cap (0,t)\}$ and put $\Omega_t = \{\omega | \, dim\, S_t(\omega) = m\}$. Then*

$$P|_{\Omega_t} \circ [X(t)]^{-1} \ll \lambda^m.$$

Proof. Denote, by S_t^n, a linear span of the set of vectors $\{\mathcal{E}_\tau^{-1} \cdot \Delta(X(\tau-), p(\tau)), \tau \in \mathcal{D} \cap (0,t), \|p(\tau)\|_{\mathbb{R}^d} \ge \frac{1}{n}\}$ and put $\Omega_t^n = \{\omega | \dim S_t^n(\omega) = m\}$. It is clear that $\Omega_t = \bigcup_{n \ge 1} \Omega_t^n$, and thus it is enough to prove that $P|_{\Omega_t^n} \circ [X(t)]^{-1} \ll \lambda^m$ for a given n.

Let n be fixed. Consider the family of differential grids $\{\mathcal{G}^N, N \in \mathbb{N}\}$ of the form

$$\Gamma_i^N = \Gamma^n \equiv \{u | \|u\| \ge \frac{1}{n}\}, \quad a_i^N = b_{i-1}^N = \frac{i-1}{N}, \quad h_i^N(s) = h(\frac{s - a_i^N}{b_i^N - a_i^N}), \quad s \in (a_i, b_i), i \in \mathbb{N},$$

where $h \in H_0$ is some function such that $Jh > 0$ inside $(0,1)$ and $Jh = 0$ outside $(0,1)$.

Our aim is to show that almost surely

$$\Omega_t^n \subset \bigcup_N \{\omega | \tilde{\Sigma}^{X(t), \mathcal{G}^N}(\omega) \text{ is non-degenerate }\}. \tag{4.5}$$

Here $\Sigma^{X(t), \mathcal{G}^N}$ is the Malliavin matrix for the random vector $X(t)$ (see Theorem 3.1). Theorem 3.1 together with (4.5) immediately imply the needed statement.

Denote $\mathcal{D}^n \equiv \mathcal{D}^{\Gamma^n}$, $A_N^{n,t} = \Big\{\omega | \mathcal{D} \cap \{\frac{i-1}{N}, i \in \mathbb{N}\} = \emptyset, \#\{\tau \in \mathcal{D}^n \cap (a_i, b_i)\} \subset \{0,1\}, i = 1, \dots, [Nt+1]\Big\}$. Since $\Gamma^n \in \Pi_{fin}$, one has that almost surely

$$\Omega_t^n \subset \bigcup_N [\Omega_t^n \cap A_N^{n,t}].$$

Thus in order to prove (4.5), it is sufficient to show that, for every N, the matrix $\Sigma^{X(t), \mathcal{G}^N}$ is non-degenerate on the set $\Omega_t^n \cap A_N^{n,t}$.

A change of the point measure outside $[0,t]$ does not change $X(t)$, thus

$$(D^{\mathcal{G}^N} X(t), l)_{\ell_2(\mathcal{G}^N)} = 0 \text{ for any } l: \quad l_i = 0, \quad i \le [Nt+1].$$

This means that the matrix $\Sigma^{X(t), \mathcal{G}^N}$ is the Grammian for the finite family of the vectors in $\mathbb{R}^m$

$$Y^1 \equiv (D^{\mathcal{G}^N} X(t), l^1)_{\ell_2(\mathcal{G}^N)}, \dots, Y^{[Nt+1]} \equiv (D^{\mathcal{G}^N} X(t), l^{[Nt+1]})_{\ell_2(\mathcal{G}^N)}, \quad l^r = (l_i^r), \quad l_i^r = \begin{cases} 1, & i = r, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore $\Sigma^{X(t), \mathcal{G}^N}$ is non-degenerate iff the family $\{Y^r, r = 1, \dots, [Nt+1]\}$ is of the maximal rank.

The family $\{Y^r\}$ on the set $A_N^{n,t}$ can be given explicitly. First of all, let us write the solution to equation (4.2) in the following form:

$$Y_h^\Gamma(t) = \mathcal{E}_t \int_0^t \int_\Gamma Jh(s)\mathcal{E}_s^{-1}\Delta(X(s-),u)\,\nu(ds,du), \quad t \geq 0, \quad h \in H_0, \Gamma \in \Pi_{fin}. \tag{4.6}$$

Taking in (4.6) $\Gamma = \Gamma^n$ and $h = h_r^N, r = 1,\ldots,[Nt+1]$, we obtain that, on the set $A_N^{n,t}$ $Y^r = c_r\mathcal{E}_t\tilde{Y}^r$, $r = 1,\ldots,[Nt+1]$, where

$$(c_r, \tilde{Y}^r) = \begin{cases} (h_r^N(\tau_r), \mathcal{E}_{\tau_r}^{-1}\cdot\Delta(X(\tau_r-), p(\tau_r))), & \{\tau \in \mathcal{D}^n \cap (a_r,b_r)\} = \{\tau_r\}, \\ (1,0), & \{\tau \in \mathcal{D}^n \cap (a_r,b_r)\} = \emptyset. \end{cases}$$

The matrix $\mathcal{E}_t$ is non-degenerate, the constants $\{c_r\}$ are positive on $A_N^{n,t}$. This means that $\{Y^r\}$ has the maximum rank iff the same holds true for $\{\tilde{Y}^r\}$. But the family $\{\tilde{Y}^r\}$ contains all the vectors

$$\{\mathcal{E}_\tau^{-1}\cdot\Delta(X(\tau-), p(\tau)), \tau \in \mathcal{D}^n \cap (0,t)\}$$

and therefore has the maximal rank on Ω_t^n. This means that $\{Y^r\}$ has the maximal rank on $\Omega_t^n \cap A_N^{n,t}$ and (4.5), together with the statement of the lemma, holds true. The lemma is proved.

Lemma 4.3. *Under the condition of Theorem 1.1,*

$$\gamma_n \equiv \inf_{x\in\bar{B}(x_*,\varepsilon_*), v\neq 0} \Pi\Big(u : \|u\| \geq \frac{1}{n}, (\Delta(x,u),v)_{\mathbb{R}^d} \neq 0\Big) \to +\infty, \quad n \to +\infty. \tag{4.7}$$

This statement follows immediately from the Dini theorem applied to the monotone sequence of lower semi-continuous functions

$$\phi_n : \bar{B}(x_*,\varepsilon_*) \times \{\|v\| = 1\} \ni (x,v) \mapsto \Pi\Big(u : \|u\| \geq \frac{1}{n}, (\Delta(x,u),v)_{\mathbb{R}^d} \neq 0\Big).$$

Proof of the Theorem 1.1 Denote by $\mathcal{S}$ the set of all proper subspaces of $\mathbb{R}^m$. This set can be parametrized in such a way that it becomes a Polish space, and, for every of the random vectors $\xi_1,\ldots,\xi_k$, the map $\omega \mapsto \mathrm{span}\,(\xi_1(\omega),\ldots,\xi_k(\omega))$ defines the random element in $\mathcal{S}$.

For every $n \geq 1$, consider the set $\mathcal{D}^n = \{\tau_1^n, \tau_2^n, \ldots\}$. For a given $S^* \in \mathcal{S}$, $\delta > 0$, let us consider the event

$$A_\delta^n = \{S_\delta^n \not\subset S^*\} = \{\exists i : \tau_i^n \leq \delta, \mathcal{E}_{\tau_i^n}^{-1}\Delta(X(\tau_i^n-), p(\tau_i^n)) \notin S^*\}$$

(see the beginning of the proof of Lemma 4.2 for the notation S_t^n). One has that $\Omega\backslash A_\delta^n \subset B_\delta \cup C_\delta^n$, where $B_\delta = \{\exists s \in [0,\delta] : X(s-) \notin \bar{B}(x_*,\varepsilon_*)\}$,

$$C_\delta^n = \bigcap_i \Big[\{\tau_i^n > \delta\} \cup \{X(\tau_i^n-) \in \bar{B}(x_*,\varepsilon_*), \mathcal{E}_{\tau_i^n}^{-1}\Delta(X(\tau_i^n-), p(\tau_i^n)) \in S^*, \tau_i^n \leq \delta\}\Big].$$

The distribution of the value $p(\tau_i^n)$ is equal to $\lambda_n^{-1}\Pi|_{\Gamma^n}$, where $\Gamma^n = \{u \| \|u\| \geq \frac{1}{n}\}, \lambda^n = \Pi(\Gamma^n)$. Moreover, this value is independent with the σ-algebra $\mathcal{F}_{\tau_i^n-}$, and, in particular, with the variables $X(\tau_i^n-), \mathcal{E}_{\tau_i^n}$. This provides the estimate

(4.8)

$$P[\{\tau_i^n > \delta\} \cup \{X(\tau_i^n-) \in \bar{B}(x_*,\varepsilon_*), \mathcal{E}_{\tau_i^n}^{-1}\Delta(X(\tau_i^n-), p(\tau_i^n)) \in S^*, \tau_i^n \leq \delta\}\Big|\mathcal{F}_{\tau_i^n-}\Big] \leq \mathbf{1}_{\{\tau_i^n > \delta\}} + (1 - \frac{\gamma_n}{\lambda_n})\mathbf{1}_{\{\tau_i^n \leq \delta\}}.$$

It follows from (4.8) that

$$P(C_\delta^n) \leq E\left(1 - \frac{\gamma_n}{\lambda_n}\right)^{\nu([0,\delta]\times\Gamma^n)} = \exp\{-\delta\gamma_n\} \to 0, n \to +\infty.$$

Since $A_\delta^n \subset \{S_\delta \not\subset S^*\}$, this means that almost surely

$$\{S_\delta \subset S^*\} \subset B_\delta. \tag{4.9}$$

Now we take $\delta < \frac{t}{m}$ and iterate (4.9) on the time intervals $[0,\delta],[\delta,2\delta],\dots,[(m-1)\delta,m\delta]$ with $S_1^*=\{0\}, S_2^*=S_\delta,\dots,S_m^*=S_{(m-1)\delta}$ (we can do this due to the Markov property of X). We obtain that

$$\{\dim S_t < m\}\subset \bigcup_{k=1}^m \{\dim S_{(k-1)\delta}=\dim S_{k\delta}<m\}\subset B_{m\delta}.$$

Since $P(B_{m\delta})\to 0, \delta\to 0+$, this provides that $P\{\dim S_t<m\}=0$, which together with Lemma 4.2 gives the needed statement. The theorem is proved.

Proof of Theorem 1.2. Due to statement II of Theorem 4.1, the solutions $X_n(x_n,t_n)$ to (1.3) are stochastically differentiable and their derivatives are given by SDEs of the form (4.2). The usual localization arguments allows us to restrict the consideration to the case where $\{a_n\}$ are uniformly bounded together with their derivatives and Π is supported by some bounded set. Then, applying Theorem 4, [10], Chapter 4.2, we obtain that, for any $p>1$, $X_n(x_n,t_n)$ converge to $X(x_*,t)$ in the L_p sense, together with their stochastic derivatives given by (4.2). This means that, for every finite differential grid $\mathcal{G}$ and any $p>1$,

$$X_n(x_n,t_n)\to X(x_*,t)\quad \text{in } W_p^1(\mathcal{G},\mathbb{R}^m),\quad n\to+\infty.$$

Thus the statement of Theorem 1.2 follows from Theorem 3.3.

4.3. **The proofs of Propositions 2.1 – 2.3.** *Proof of the Proposition 2.1.* Take $\varepsilon_*=\frac{\delta_*}{2}$. Then, for every $x\in\bar B(x_*,\varepsilon_*)$,

$$\{u|\Delta(x,u)=0\}=\{u|a(x+u)=a(x)\}\subset\{|u|>C_\bullet\,\delta_*\}\cup\{u|x+u\in N(a,a(x))\cap(x_*-\delta_*,x_*+\delta_*)\}=\Delta_1\cup\Delta_2.$$

Here we used that a is Lipschitz. The set Δ_2 is finite and therefore $\Pi(\Delta_2)<+\infty$. The set Δ_1 is separated from 0 and therefore $\Pi(\Delta_1)<+\infty$. Since $\Pi(\mathbb{R})=+\infty$, this means that $\Pi(\Delta(x,u)\neq 0)=+\infty$. Proposition is proved.

The proof of the Proposition 2.2 is almost trivial: for a given $x\in B(x_*,\varepsilon_*), v\in S_m$ one should take $w=w(x,v)$, given by the Definition 1.3, and for this v choose $\varrho\in(0,1)$ such that $\Pi(V(w,\varrho))=+\infty$ (this is possible since Π satisfies the wide cone condition). Then, for every $D>0$, $\Pi(u\in V(w,\varrho),\|u\|\le D)=+\infty$, and (1.2) follows from (1.1).

Proof of the Proposition 2.3. Consider the set Φ_{x_*,ε_*} of the functions $\phi_x:\mathbb{R}^m\ni u\mapsto a(x+u)-a(x)\in\mathbb{R}^m, x\in B(x_*,\varepsilon_*)$. It is easy to see that if for every linear subspace $L_v\equiv\{y|(y,v)=0\}, v\in S_m$,

$$\Pi(u|u\notin\phi^{-1}(L_v))=+\infty,\quad \phi\in\Phi_{x_*,\varepsilon_*},$$

then (1.2) holds true. In the case **b**, Φ_{x_*,ε_*} contains the unique function $\phi(u)=Au$. Since A is non-degenerate, $\phi^{-1}(L_v)$ is a proper linear subspace of $\mathbb{R}^m$ for every $v\in S_m$, and (2.2) provides (1.2). In the case **a**, $\phi_x\in C^1(\mathbb{R}^m,\mathbb{R}^m)$, and for ε_* small enough $\det\nabla\phi_x(0)=\det\nabla a(x)\neq 0, x\in B(x_*,\varepsilon_*)$. Then $\phi_x^{-1}(L_v)$ is a proper smooth subspace of $\mathbb{R}^m$ for every $v\in S_m$, and (2.1) provides (1.2). Proposition is proved.

5. Smoothness of the density of the solution to the Cauchy problem

5.1. **The irregularity properties of the density.** We start our exposition with the easier part: the proof of Theorem 1.4 and Proposition 2.4, that give the irregularity properties of $p_{x,t}$.

Recall that the function a is supposed to be globally Lipschitz and the jump noise is supposed to satisfy the moment condition (1.4).

Proof of Theorem 1.4: the case $m=1$. For $\varepsilon\in(0,1)$, denote $M^\varepsilon=\int_{\varepsilon<|u|\le 1}u\,\Pi(du)$ and consider a decomposition of the process U_t of the form

$$U_t=U_0+R_t^\varepsilon+V_t^\varepsilon-tM^\varepsilon,\quad R_t^\varepsilon=\int_0^t\int_{\|u\|\le\varepsilon}u\tilde\nu(ds,du),\quad V_t^\varepsilon=\int_0^t\int_{\|u\|>\varepsilon}u\nu(ds,du).$$

R_t^ε is a martingale, and its quadratic variation is equal to

$$[R^\varepsilon]_t = \sum_{s\le t}(R_s - R_{s-})^2 = \int_0^t \int_{\|u\|\le\varepsilon} u^2\nu(ds,du).$$

We have that

$$E[R^\varepsilon]_t^2 = E\left[\int_0^t \int_{\|u\|\le\varepsilon} u^2\tilde\nu(ds,du)\right]^2 + \left[\int_0^t \int_{\|u\|\le\varepsilon} u^2\Pi(du)\,ds\right]^2 =$$
$$= t\int_{\|u\|\le\varepsilon} u^4\Pi(du) + t^2\left[\int_{\|u\|\le\varepsilon} u^2\Pi(du)\right]^2,$$

and therefore, for ε small enough,

$$E[R^\varepsilon]_t^2 \le C_\bullet\left[\varepsilon^2\ln\frac{1}{\varepsilon}\right]\rho(\varepsilon).$$

Applying the Chebyshev and Burkholder inequalities, we obtain that, for every given $\alpha>0$,

$$P(\sup_{s\le t}|R_s^\varepsilon|\ge\varepsilon^{1-\alpha}) \le \frac{C_\bullet\, E[R^\varepsilon]_t^2}{\varepsilon^{2-2\alpha}} \le C_\bullet\left[\varepsilon^{2\alpha}\ln\frac{1}{\varepsilon}\right]\rho(\varepsilon).$$

Next, for every $\varepsilon\in(0,1)$

$$P(V_s^\varepsilon=0, s\in[0,t]) = \exp[-t\Pi(|u|>\varepsilon)] \ge \exp[-\varepsilon^{-2}t\int_{\mathbb{R}} u^2\wedge\varepsilon^2\Pi(du)] = \exp[t\rho(\varepsilon)\ln\varepsilon] = \varepsilon^{t\rho(\varepsilon)}.$$

Denote $A_\alpha^\varepsilon = \{|R_s^{\varepsilon_n}|\le\varepsilon_n^{1-\alpha}, V_s^\varepsilon=0, s\in[0,t]\}$. Since R^ε, V^ε are independent, we have

$$P(A_\alpha^\varepsilon) \ge \varepsilon^{t\rho(\varepsilon)}\left[1 - C_\bullet\left[\varepsilon^{2\alpha}\ln\frac{1}{\varepsilon}\right]\rho(\varepsilon)\right].$$

Considering a sequence $\varepsilon_n\to 0+$ such that $\rho(\varepsilon_n)\to\boldsymbol{\rho}, n\to+\infty$, we obtain that, for n big enough,

$$P(A_\alpha^{\varepsilon_n}) \ge \frac{1}{2}\varepsilon_n^{t(\boldsymbol{\rho}+\alpha)}.$$

Denote, by $X^n(x,t)$, the solution to the ODE

$$X^n(x,t) = x + \int_0^t a(X^n(x,s))\,ds - tM^{\varepsilon_n}. \tag{5.1}$$

By the construction of the set $A_\alpha^{\varepsilon_n}$, we have that on this set

$$|X(x,s)-X^n(x,s)| \le L\int_0^s |X(x,r)-X^n(x,r)|dr + \varepsilon^{1-\alpha}, \quad s\in[0,t],$$

where L denotes the Lipschitz constant for a. Then, by the Gronwall lemma, $|X(x,t)-X^n(x,t)|\le e^{Lt}\varepsilon_n^{1-\alpha}$ on the set $A_\alpha^{\varepsilon_n}$. Thus there exist two sequences $y_n = X^n(x,t)-e^{Lt}\varepsilon_n^{1-\alpha}$, $z_n = X^n(x,t)+e^{Lt}\varepsilon_n^{1-\alpha}$ such that, for n big enough,

$$P(y_n\le X(x,t)\le z_n) \ge C_\bullet\,(z_n-y_n)^{(\boldsymbol{\rho}+\alpha)\cdot\frac{t}{1-\alpha}}. \tag{5.2}$$

Now we can complete the proof. For $y<z$

$$\int_y^z f(v)\,dv \le \|f\|_{L_\infty}(z-y), \quad \int_y^z f(v)\,dv \le \|f\|_{L_r}\left[\int_y^z 1^{\frac{r}{r-1}}\,dv\right]^{\frac{r-1}{r}} = \|f\|_{L_r}(z-y)^{\frac{r-1}{r}}, r\in[1,+\infty). \tag{5.3}$$

Let $t\boldsymbol{\rho} < 1-\frac{1}{r}$. Then there exists $\alpha>0$ such that $(\boldsymbol{\rho}+\alpha)\cdot\frac{t}{1-\alpha} < 1-\frac{1}{r}$ and (5.2) together with (5.3) indicates that $p_{x,t}\notin L_r(\mathbb{R})$. This proves the statement **a1**. Analogously, if $\boldsymbol{\rho}t<1$, then there exists $\alpha>0$ such that $(\boldsymbol{\rho}+\alpha)\cdot\frac{t}{1-\alpha} < 1$ and (5.2), (5.3) indicate that $p_{x,t}$ is not bounded, i.e. $p_{x,t}\notin CB^0(\mathbb{R})$. This proves the statement **b1**. Under condition (1.4) there exists $\lim_{n\to+\infty}M^{\varepsilon_n} = M^0$ and the sequences $\{y_n\},\{z_n\}$ are bounded, that implies statements **a**,**b**.

Proof of Theorem 1.4: the case $m>1$. Consider a decomposition of the process $U=(u^1,\dots,U^m)$ of the form

$$U^i_t=U^i_0+R^{\varepsilon,i}_t+V^{\varepsilon,i}_t-tM^{\varepsilon,i},\quad i=1,\dots,m,$$

where $R^{\varepsilon,i}_t=\int_0^t\int_{|u^i|\le\varepsilon}u^i\tilde\nu(ds,du), V^{\varepsilon,i}_t=\int_0^t\int_{|u^i|>\varepsilon}u^i\nu(ds,du), M^{\varepsilon,i}=\int_{\varepsilon<|u^i|,\|u\|\le1}u^i\,\Pi(du)$. Then, analogously to the proof of the case $m=1$, one can verify that

$$P(\sup_{s\le t}\|R^\varepsilon_s\|\ge\varepsilon^{1-\alpha})\le C_\bullet\left[\varepsilon^{2\alpha}\ln\frac{1}{\varepsilon}\right]\vartheta(\varepsilon).$$

On the other hand,

$$P(V^\varepsilon_s=0,s\in[0,t])=\exp\left[-t\Pi\left(\bigcup_{i=1}^m\{|u^i|>\varepsilon\}\right)\right]\ge\exp\left[-t\sum_{i=1}^m\Pi\left(\{|u^i|>\varepsilon\}\right)\right]\ge\varepsilon^{mt\vartheta(\varepsilon)}.$$

Then, just as in the case $m=1$, for every $\alpha\in(0,1)$ there exist sequences $\{y^i_n\}\subset\mathbb{R}^m, i=1,\dots,m$ and $\{\delta_n\}\mathbb{R}^+$ such that $\delta_n\to0$ and

$$P(X^i_n\in[y^i_n,y^i_n+\delta_n],i=1,\dots,m)\ge C_\bullet\,\delta_n^{(\boldsymbol{\vartheta}+\alpha)\cdot\frac{mt}{1-\alpha}}. \tag{5.4}$$

The arguments analogous to those used in the proof of the case $m=1$ show that (5.4) implies statements **a**,**b**,**a1**,**b1** of Theorem 1.4. The theorem is proved.

Proof of Proposition 2.4. If $\boldsymbol{\rho}_1=+\infty$, then the statement is trivial. Thus we consider only the case $\boldsymbol{\rho}_1<+\infty$. Without losing generality, we can suppose that $\Pi((-\infty,0))=0$.

Consider a sequence $\{\varepsilon_n\}$ such that $\rho_1(\varepsilon_n)\to\boldsymbol{\rho}_1$. Since $\rho_1(\varepsilon)\ge\rho_2(\varepsilon)$, for the sequences $\{y_n\},\{z_n\}$ given in the proof of Theorem 1.4 (the case $m=1$), the following estimate holds true:

$$P(y_n\le X(x,t)\le z_n)\ge C_\bullet\,(z_n-y_n)^{(\boldsymbol{\rho}_1+\alpha)\cdot\frac{t}{1-\alpha}}. \tag{5.5}$$

Denote, by $X^*(x,t)$, the solution to an ODE of the type (5.1) with M^{ε_n} replaced by M^0. It follows from the comparison theorem that the law of $X(x,t)$ is supported by $[X^*(x,t),-\infty)$ and the density $p_{x,t}$ is equal to zero on $(-\infty,X^*(x,t))$. On the other hand, $M^0-M^{\varepsilon_n}\le\left[\varepsilon_n\ln\frac{1}{\varepsilon_n}\right]\rho_1(\varepsilon_n)=o(\varepsilon_n^{1-\alpha}),n\to+\infty$ and therefore, for n big enough, $(y_n,z_n)\cap(-\infty,X^*(x,t))\ne\emptyset$. Therefore one can show iteratively that if $p_{x,t}\in CB^k$, then

$$\left|\frac{d^{k-1}p_{x,t}}{dy^{k-1}}(y)\right|\le C_\bullet\,(z_n-y_n),\left|\frac{d^{k-2}p_{x,t}}{dy^{k-2}}(y)\right|\le C_\bullet\,(z_n-y_n)^2,\dots,|p_{x,t}(y)|\le C_\bullet\,(z_n-y_n)^k,\quad y\in(y_n,z_n),$$

and $P(y_n\le X(x,t)\le z_n)\le C_\bullet\,(z_n-y_n)^{k+1}$. Comparing this estimate with (5.5) and taking α sufficiently small, we obtain the needed statement. The proposition is proved.

5.2. **Smoothness of the density.** The crucial difficulty in the proof of the smoothness of the density is that the stochastic derivative Y^Γ_h of the variable $X(x,t)$, given by Theorem 4.1, is not stochastically differentiable w.r.t. $\{T^\Gamma_{rh}\}$. This formally does not allow one to apply the standard Malliavin-type regularity results. Moreover, the detailed analysis shows that this difficulty is not only formal and the integration-by-parts formula for the functionals of $X(x,t)$ (formula (5.24) below) actually contains some additional "singular" terms. Below we introduce the calculus of variations based on such integration-by-parts formula and obtain the sufficient conditions for the density of the law of the solution to (0.1) to be smooth.

Let us introduce some necessary constructions. We would like to have an opportunity to divide any "portion of the jump mass" Π into an arbitrary number of parts. Such an opportunity is guaranteed by the following construction: we suppose that the point measure ν, correspondent to the process U, is in fact a projection of another point measure $\boldsymbol{\nu}$ with a more wide phase space and the specially constructed Lévy measure $\boldsymbol{\Pi}$.

To be precise, we suppose that the probability space is generated by a Poisson random point measure $\boldsymbol{\nu}$ on $\mathbb{R}^+ \times \mathbb{R}^{m+1}$ with the intensity measure $\lambda^1 \times \mathbf{\Pi}$, $\mathbf{\Pi} \equiv \Pi \times (\lambda^1|_{[0,1]})$, and ν is expressed through $\boldsymbol{\nu}$ by

$$\nu([0,t] \times \Gamma) = \boldsymbol{\nu}([0,t] \times \Gamma \times [0,1]), \quad t \in \mathbb{R}^+, \Gamma \in \Pi_{fin}.$$

It is easy to see that such supposition does not restrict generality, since for a given ν we can construct $\boldsymbol{\nu}$, making an appropriate extension of the initial probability space.

For the "extended" random point measure $\boldsymbol{\nu}$, we will use the terminology and constructions from Section 3. Further we denote $\mathbf{u} = (u, y) \in \mathbb{R}^{m+1}$, the subsets of $\mathbb{R}^{m+1}$ are denoted by bold symbols, such as $\mathbf{\Gamma}$. We also denote, by $\mathbf{p}(\cdot) = (p(\cdot), q(\cdot))$, the point process corresponding to $\boldsymbol{\nu}$.

Given the measure Π, let us construct the monotonously decreasing sequence $\{\varepsilon_n, n \in \mathbb{Z}\}$ in the following way:

$$\varepsilon_0 = 1, \quad \frac{\varepsilon_{n+1}}{\varepsilon_n} = 1 - (|n| + 2)^{-1}, n \in \mathbb{Z}.$$

By the construction, the sequence $\{\varepsilon_n\}$ has the following properties:

$$\varepsilon_n \downarrow 0, n \to +\infty, \quad \varepsilon_n \to \infty, n \to -\infty, \quad \frac{\varepsilon_{n+1}}{\varepsilon_n} \to 1, n \to \infty, \quad \sup_n \frac{\varepsilon_n}{\varepsilon_{n+1}} \leq 2.$$

Denote $I_n \equiv \{u | \|u\| \in [\varepsilon_{n+1}, \varepsilon_n)\}$. Let $t \in \mathbb{R}^+, \gamma \in (0, \frac{1}{2}), B > 0$ be fixed, define the numbers $K_n \in \mathbb{N}, n \in \mathbb{Z}$ by

$$K_n = \left[\max\left(B, 2t\Pi(I_n), \frac{3}{\gamma} \cdot 2^{|n|-2} t^2 \Pi(I_n)\right)\right] + 2,$$

where $[x] \equiv \max\{k \in \mathbb{Z}, k \leq x\}$. By the construction,

$$K_n > B, \quad \frac{t}{K_n}\Pi(I_n) < \frac{1}{2}, \quad \frac{t^2}{K_n}\Pi^2(I_n) < \frac{2\gamma}{3} \cdot 2^{-|n|}.$$

We consider all the sets of the type $I_n \times [\frac{k-1}{K_n}, \frac{k}{K_n}) \subset \mathbb{R}^{m+1}, k = 1, \dots, K_n, n \in \mathbb{Z}$, and enumerate them in an arbitrary way by the parameter $i \in \mathbb{N}$. The i-th set from this family will be denoted by $\mathbf{\Gamma}_i^\gamma$. Now, we can consider the the grid $\mathcal{G}^\gamma$ for the random point measure $\boldsymbol{\nu}$ in the following way.

1) Every time interval $[a_i^\gamma, b_i^\gamma)$ is equal to $[0, t)$.

2) The family of sets $\{\mathbf{\Gamma}_i^\gamma\}$ is the one constructed before.

3) For every i, the function h_i^γ has the form $(\varepsilon_n^{-1} \wedge 1)h$, where $n = n(i)$ is such that $\mathbf{\Gamma}_i^\gamma = I_n \times [\frac{k-1}{K_n}, \frac{k}{K_n})$ for some k. The function $h \in C^\infty(\mathbb{R})$ is such that $Jh = 0$ outside $(0, t)$, $Jh > 0$ inside $(0, t)$ and $Jh = 1$ on $(\beta, t - \beta)$, where the constant $\beta \in (0, \frac{1}{2})$ will be determined later on.

Denote $\Xi^\gamma = \bigcap_i \{\#\{\tau \in \mathcal{D} \cap [0, 1) | \mathbf{p}(\tau) \in \mathbf{\Gamma}_i^\gamma\} \leq 1\}$. All the variables $\#\{\tau \in \mathcal{D} \cap [0, 1) | \mathbf{p}(\tau) \in \mathbf{\Gamma}_i^\gamma\}$ are independent Poissonian variables with the intensities $\lambda_i \equiv \frac{t}{K_{n(i)}}\Pi(I_{n(i)})$. For any Poissonian variable ξ with the intensity λ, the inequality $P(\xi > 1) \leq \frac{\lambda^2}{2}$ holds true. Thus

(5.6)

$$P(\Xi^\gamma) \geq \prod_{n \in \mathbb{Z}} \prod_{k=1}^{K_n} \left(1 - \frac{t^2 \Pi^2(I_{n(i)})}{2K_{n(i)}^2}\right) \geq 1 - \sum_{n \in \mathbb{Z}} \sum_{k=1}^{K_n} \frac{t^2 \Pi^2(I_{n(i)})}{2K_{n(i)}^2} = 1 - \sum_{n \in \mathbb{Z}} \frac{t^2 \Pi^2(I_{n(i)})}{2K_{n(i)}} > 1 - \sum_{n \in \mathbb{Z}} \frac{\gamma}{3} \cdot 2^{-|n|} = 1 - \gamma.$$

Our trick is to replace the initial probability P by

$$P^\gamma(\cdot) = P(\cdot | \Xi^\gamma) = \frac{P(\cdot \cap \Xi^\gamma)}{P(\Xi^\gamma)}.$$

We will study firstly the distribution of $X(x, t)$ w.r.t. P^γ and then tend γ to 0. The key point here is the following analog of the classical Fourier lemma (see [26] or Lemma 8.1 [11]). Below we denote, by E^γ, the expectation w.r.t. P^γ.

Lemma 5.1. *Suppose that, for some $k \geq 0$, there exists constants $\mathsf{C}_1, \dots, \mathsf{C}_{k+m} \in \mathbb{R}^+$ such that, for every $\gamma \in (0, \frac{1}{2}), F \in C_b^\infty(\mathbb{R}^m), n \leq k+m, \alpha_1, \dots, \alpha_n \in \{1, \dots, m\}$,*

$$\left| E^\gamma \Big[\frac{\partial}{\partial x_{\alpha_1}} \dots \frac{\partial}{\partial x_{\alpha_n}} F \Big] (X(t)) \right| \leq \mathsf{C}_n \sup_x |F(x)|. \tag{5.7}$$

Then $P(X(t) \in dx) = p(x)dx$ with $p \in CB^k(\mathbb{R}^m)$.

Proof. The Fourier lemma provides that $P^\gamma(X(t) \in dx) = p^\gamma(x)dx$ with $p^\gamma \in CB^k(\mathbb{R}^m)$ and

$$\left\| \frac{\partial}{\partial x_{\alpha_1}} \dots \frac{\partial}{\partial x_{\alpha_k}} p^\gamma \right\|_{L_\infty} \leq \mathsf{C}_m, \quad \alpha_1, \dots, \alpha_k \in \{1, \dots, m\}.$$

Due to (5.6), the measures $P^\gamma(X(t) \in \cdot)$ weakly converge to $P(X(t) \in \cdot)$, $\gamma \to 0+$. This implies the needed statement. The lemma is proved.

Thus, our further goal is to construct the grids $\mathcal{G}^\gamma$ in the special way in order to provide (5.7) to hold true. Let us mention that Ξ^γ is invariant w.r.t. $T_{rh_i^\gamma}^{\mathbf{\Gamma}_i^\gamma}$, and $\mathbb{1}_{\Xi^\gamma} \in W_p^1(\mathcal{G}^\gamma, \mathbb{R})$ with $D_p^{\mathcal{G}^\gamma} \mathbb{1}_{\Xi^\gamma} = 0$, $p \in (1, +\infty)$. This means that the "censoring" operation $P \mapsto P^\gamma$ described above is adjusted with the differential structure. On the other hand, the following proposition shows that P^γ is some kind of a mixture of the Bernoulli and uniform distributions. Such a measure appears to be more convenient for us to deal with, than the initial Poisson one. Below we omit the superscript γ in the notation for Ξ^γ and $\mathbf{\Gamma}_i^\gamma$ (but not for P^γ).

Proposition 5.1. *Denote*

$$\Xi_i^0 = \{\{\tau \in \mathcal{D} \cap [0,1) | \mathbf{p}(\tau) \in \mathbf{\Gamma}_i\} = \emptyset\}, \quad \Xi_i^1 = \Xi \backslash \Xi_i^0 = \{\{\tau \in \mathcal{D} \cap [0,1) | \mathbf{p}(\tau) \in \mathbf{\Gamma}_i\} = \{\tau_i\}\}.$$

Then

a) $P^\gamma(\Xi_i^0) = \frac{1}{1+\lambda_i}, P^\gamma(\Xi_i^1) = \frac{\lambda_i}{1+\lambda_i}$;

b) the distribution of τ_i w.r.t. $P(\cdot|\Xi_i^1) = P^\gamma(\cdot|\Xi_i^1)$ coincides with the uniform distribution on $[0,t]$ (below we denote this distribution by λ_t^1);

c) the distribution of $\mathbf{p}(\tau_i)$ w.r.t. $P(\cdot|\Xi_i^1)$ is equal to $\mu_i(\cdot) = \frac{\mathbf{\Pi}(\cdot \cap \mathbf{\Gamma}_i)}{\mathbf{\Pi}(\mathbf{\Gamma}_i)}$;

d) for any $i_1, \dots, i_k \in \mathbb{N}, i_j \neq i_l, j \neq l$, $a_1, \dots, a_k \in \{0,1\}$ the sets $\Xi_{i_1}^{a_1}, \dots, \Xi_{i_k}^{a_k}$ are jointly independent w.r.t. P^γ;

e) for any $i_1, \dots, i_k \in \mathbb{N}, i_j \neq i_l, j \neq l$, the variables $\tau_{i_1}, \dots, \tau_{i_k}, \mathbf{p}(\tau_{i_1}), \dots, \mathbf{p}(\tau_{i_k})$ are jointly independent w.r.t. $P^\gamma(\cdot | \bigcap_{j=1}^k \Xi_{i_j}^1)$.

Proof. Denote by $\boldsymbol{\nu}_i, i \in \mathbb{N}$ the point measures, defined on $\mathbb{R}^+ \times \mathbb{R}^{m+1}$ by

$$\boldsymbol{\nu}_i([0,s] \times \Delta) = \boldsymbol{\nu}([0,s] \times (\mathbf{\Gamma}_i \cap \Delta)), \quad s \in \mathbb{R}^+, \Delta \in \mathcal{B}(\mathbb{R}^{m+1}), i \in \mathbb{N}.$$

The following facts (valid for any disjoint family of the sets $\mathbf{\Gamma}_i, i \in \mathbb{N}$ with $\mathbf{\Pi}(\mathbf{\Gamma}_i) < +\infty$) are well known in the theory of the Lévy processes:

(i) the measures $\{\boldsymbol{\nu}_i, i \in \mathbb{N}\}$ are jointly independent;

(ii) for every $i \in \mathbb{N}$, the domain of the point process $\mathbf{p}_i$, correspondent to $\boldsymbol{\nu}_i$, is a.s. locally finite;

(iii) for every $i \in \mathbb{N}$ the sequences $\{\tau_1^i, \tau_2^i, \dots\}$ and $\{\xi_1^i, \xi_2^i, \dots\}$ of the points of the domain of $\mathbf{p}_i$ (enumerated increasingly) and correspondent values of $\mathbf{p}_i$ are independent;

(iv) the process $N_s^i \equiv \#\{k | \tau_k \leq s\}$ is a Poisson process with the intensity $\mathbf{\Pi}(\mathbf{\Gamma}_i)$;

(v) $\{\xi_k^i, k \geq 1\}$ are i.i.d. random vectors in $\mathbb{R}^{m+1}$ with their common distribution equal to $\frac{\mathbf{\Pi}(\cdot \cap \mathbf{\Gamma}_i)}{\mathbf{\Pi}(\mathbf{\Gamma}_i)}$.

For any $i \in \mathbb{N}$ the sets $\Xi_i^a, a = 1,2$ belong to $\sigma(N_\cdot^i)$, and $\Xi \equiv \Xi^\gamma = \bigcap_{i \in \mathbb{N}} [\Xi_i^0 \cup \Xi_i^1]$. Using this, one can easily verify that (i) – (v) imply statements c),d),e). For a Poisson process N with the intensity λ, we have that

$$P(N_t = 0) = e^{-t\lambda}, \quad P(N_t = 1) = (t\lambda)e^{-t\lambda} \Longrightarrow P(N_t = 0 | N_t \leq 1) = \frac{1}{1+t\lambda}, \quad P(N_t = 1 | N_t \leq 1) = \frac{t\lambda}{1+t\lambda}.$$

This provides the statement a). At last, for the moment τ of the first jump of the process N, the following relation holds:

$$P(\tau \le s|N_t = 1) = \left[(t\lambda)e^{-t\lambda}\right]^{-1} P(N_s = 1, N_t = 1) = \left[(t\lambda)e^{-t\lambda}\right]^{-1} \left\{(s\lambda)e^{-s\lambda} \cdot e^{-(t-s)\lambda}\right\} = \frac{s}{t}, \quad s \in [0,t].$$

This provides the statement b). The proposition is proved.

Consider the space $\varOmega = \prod_{i\in\mathbb{N}}(\{0,1\} \times [0,t] \times \mathbb{R}^{m+1})$ with the measure $M = \prod_{i\in\mathbb{N}}\Big(\mathrm{Be}(\frac{\lambda_i}{1+\lambda_i}) \times \lambda_t^1 \times \mu_i\Big)$, here $\mathrm{Be}(p)$ denotes the Bernoulli distribution with $P(1) = p$. For every $\varpi = (\theta_i, s_i, \mathbf{u}_i, i \in \mathbb{N}) \in \varOmega$, we define the configuration $\omega = \omega(\varpi)$ in the following way: it consists of the points $\{(s_i, \mathbf{u}_i) \in [0,t] \times \mathbb{R}^{m+1}, i \in I^1\}$, where $I^1 = \{i|\theta_i = 1\}$. Let the function $f \in L_0(\Omega, \mathcal{F}, P)$ depend only on the values of the point measure on $[0,t] \times \mathbb{R}^{m+1}$, define $\tilde f(\varpi) = f(\omega(\varpi))$. Since $P^\gamma \ll P$, Proposition 5.1 implies that the map $f \to \tilde f$ is well defined, i.e. taking a P-modification of f we obtain the function that is M-a.s. equal to $\tilde f$. Further we omit the sign $\tilde{}$ and denote by f both the function defined on Ω and its image defined on $\varOmega$.

Denote $\varOmega_i^j = \{\theta_i = j\}, j = 0,1$ and $M_i^j(\cdot) = M(\cdot|\varOmega_i^j), j = 0,1$. Denote

$$\mathsf{E}f = \int_\varOmega f(\varpi)\, M(d\varpi), \quad \mathsf{E}_i^j f = \int_\varOmega f(\varpi)\, M_i^j(d\varpi), \quad j = 0,1.$$

Define the transformation $\varepsilon_i^{s,\mathbf{u}} : \varOmega \to \varOmega_i^1, (s,\mathbf{u}) \in [0,t] \times \mathbf{\Gamma}_i$ in the following way: it does not change all coordinates with indices not equal to i and replaces $(\theta_i, s_i, \mathbf{u}_i)$ by $(1, s, \mathbf{u})$. The restriction of this operator on $\varOmega_i^0$ is just an appropriate version of the operator $\varepsilon^+_{(s,\mathbf{u})}$ adding the point $(s,\mathbf{u})$ to the configuration (see [29]). Denote, by the same symbol $\varepsilon_i^{s,\mathbf{u}}$, the transformation

$$L_0(\Omega, P) \ni f(\cdot) \mapsto f(\varepsilon_i^{s,\mathbf{u}} \cdot) \in L_0(\Omega_i^0, M_i^0).$$

Recall (see the discussion in [29], Section 1) that, for two different modifications f_1, f_2 of $f \in L_0(\Omega, P)$, the functions $\varepsilon_i^{s,\mathbf{u}} f_1, \varepsilon_i^{s,\mathbf{u}} f_2$ may be not equal to M_i^0 a.s. for the given $(s,\mathbf{u})$. But the set $\{(s,\mathbf{u}) : \varepsilon_i^{s,\mathbf{u}} f_1 \ne \varepsilon_i^{s,\mathbf{u}} f_2\}$ has zero $\lambda_t^1 \times \mu_i$-measure. This means that the family of the transformations $\{\varepsilon_i^{s,\mathbf{u}}, (s,\mathbf{u}) \in [0,t] \times \mathbf{\Gamma}_i\}$ is well defined in the $L_0([0,t] \times \mathbf{\Gamma}_i, \lambda_t^1 \times \mu_i)$ sense.

The following formula is a simple corollary of Proposition 5.1 and is, in fact, the main purpose of the construction given above.

Proposition 5.2. *For any $f \in L_1(\Omega, P), i \in \mathbb{N}$,*

$$\mathsf{E}_i^1 f = \frac{1}{t}\int_0^t \int_{\mathbf{\Gamma}_i} \left[\mathsf{E}_i^0 \varepsilon_i^{s,\mathbf{u}} f\right] \mu_i(d\mathbf{u}) ds. \tag{5.8}$$

Now we are going to proceed with the proof of Theorem 1.3. We will do this in two steps.

Proof of Theorem 1.3: the case $m = 1$.

Consider the functionals $f = X(t)\mathbf{1}_\Xi$ (we omit the initial value x in the notation for $X(x,t)$) and $g_i = D_{h_i}^{\mathbf{\Gamma}_i} f, i \in \mathbb{N}$. The latter derivative exists since $D_{h_i}^{\mathbf{\Gamma}_i}\mathbf{1}_\Xi = 0$. Due to Theorem 4.1, one has

$$g_i = Jh_i(\tau_i)\mathcal{E}_{\tau_i}^t\Big[a\Big(X(\tau_i-) + p(\tau_i)\Big) - a\Big(X(\tau_i-)\Big)\Big]\mathbf{1}_{\Xi_i^1}, \quad i \in \mathbb{N}.$$

Since ∇a is bounded, $|\mathcal{E}_{\tau_i}^t| \le C_\bullet$ and $\Big|a\Big(X(\tau_i-) + p(\tau_i)\Big) - a\Big(X(\tau_i-)\Big)\Big| \le C_\bullet\, |p(\tau_i)|$. We recall that $Jh_i = (\varepsilon_{n(i)}^{-1} \wedge 1)Jh$ and $\|Jh\|_\infty < +\infty$, thus

$$\sum_{i\in\mathbb{N}} g_i^2 \le C_\bullet \sum_{i\in\mathbb{N}} p_1^2(\tau_i)(\varepsilon_{n(i)}^{-2} \wedge 1)\mathbf{1}_{\Xi_i^1} \le C_\bullet \sum_{i\in\mathbb{N}}(1 \wedge \varepsilon_{n(i)}^2)\mathbf{1}_{\Xi_i^1}. \tag{5.9}$$

We have

$$E\sum_{i\in\mathbb{N}}(1 \wedge \varepsilon_{n(i)}^2)\mathbf{1}_{\Xi_i^1} = P(\Xi)\sum_{i\in\mathbb{N}}(1 \wedge \varepsilon_{n(i)}^2)\frac{\lambda_i}{1+\lambda_i} < \sum_{i\in\mathbb{N}}(1 \wedge \varepsilon_{n(i)}^2)\lambda_i =$$

$$= \sum_{n\in\mathbb{Z}}\sum_{k=1}^{K_n}(1\wedge\varepsilon_n^2)\frac{t\Pi(I_n)}{K_n} = t\sum_{n\in\mathbb{Z}}(1\wedge\varepsilon_n^2)\Pi(I_n) \le t2^2\int_{\mathbb{R}}(u^2\wedge 1)\Pi(du) < +\infty.$$

Here we used that $\varepsilon_n \le 2\varepsilon_{n+1} \le |u|$ for $u\in I_n$. Thus the series on the right-hand side of (5.9) converges in the L_1 sense, $g=(g_i)\in L_2(\Omega,P,\ell_2)$ and $f\in W_2^1(\mathcal{G}^\gamma)$ with $D^{\mathcal{G}}f=g$.

We put $Z=\|g\|^2_{\ell_2}\ge 0$. For any function $F\in C_b^\infty$, one has

$$(Z+c)^{-1}(D^{\mathcal{G}^\gamma}F(f),g)_{\ell_2} = \frac{\sum_{i\in\mathbb{N}}F'(f)g_i^2}{\sum_{i\in\mathbb{N}}g_i^2+c} \to F'(f)\mathbb{1}_{\{Z>0\}},\quad c\to 0+$$

almost surely and in every L_p. We will show below that $\{Z>0\}=\Xi$ almost surely. Thus, in order to estimate $E^\gamma F'(f) = EF'(f)\mathbb{1}_{\Xi^\gamma}$, it is enough to estimate $EF'(f)\frac{g_i^2}{Z+c}$ in such a way that is uniform in c and allows the summation over i. The key point here is the following moment estimate. For a given $k\in\mathbb{N}, i_1,\dots,i_k\in\mathbb{N}, \mathbf{u}_1,\dots\mathbf{u}_k\in\mathbb{R}^2, s_1,\dots,s_k\in[0,t]$, we denote

$$\mathsf{E}^0_{i_1,\dots,i_k}[\cdot]=\mathsf{E}[\cdot|\theta_{i_1}=\dots\theta_{i_k}=0],\quad Z^{\mathbf{u}_1,\dots,\mathbf{u}_k}_{i_1,\dots,i_k}(s_1,\dots,s_k)=\varepsilon^{s_1,\mathbf{u}_1}_{i_1}\dots\varepsilon^{s_k,\mathbf{u}_k}_{i_k}\left[\sum_{i\ne i_1,\dots,i_k}g_i^2\right]. \tag{5.10}$$

Lemma 5.2. *Let $a\in\mathbf{K}_r$ and $\frac{t}{2r}\frac{e-1}{e}\boldsymbol{\rho}_{2r}>\alpha$ for some $r\in\mathbb{N},\alpha\in[0,+\infty)$. Then, for every $k\in\mathbb{N}$, there exists $\delta>0$ such that, under an appropriate choice of the constants B,β in the construction of the grids $\mathcal{G}^\gamma$,*

$$\sup_\gamma\sup_{l\le k}\sup_{i_1,\dots,i_l\in\mathbb{N}}\sup_{\mathbf{u}_1\in\boldsymbol{\Gamma}_{i_1},\dots,\mathbf{u}_l\in\boldsymbol{\Gamma}_{i_l}}\sup_{s_1,\dots,s_l\in[0,t]}\left[\mathsf{E}^0_{i_1,\dots,i_l}[Z^{\mathbf{u}_1,\dots,\mathbf{u}_l}_{i_1,\dots,i_l}(s_1,\dots,s_l)]^{-\alpha-\delta}\right]<+\infty. \tag{5.11}$$

Proof. In order to shorten the notation, we consider only the case $k=1$, the general case is completely analogous (namely, the only change in the proof will be that the term $B-1$ in (5.17) should be replaced by $B-k$). Everywhere in the proof of the lemma, we omit the subscript near $i,\mathbf{u},s$.

We use the arguments that are not the simplest possible here, but appear to be appropriate both for the case $m=1$, and for the general case considered in Lemma 5.5 below. We return from the "censored" probability space (Ω^0_i,M^0_i) to the initial one (Ω,P) and provide (5.11) by the arguments analogous to those used in the proof of Theorem 1.1.

We have $P(\Xi^0_i)=P(\Xi)\frac{1}{1+\lambda_i}\ge C_\bullet>0$, and thus $\mathsf{E}^0_i[\cdot]=[P(\Xi^0_i)]^{-1}E[\cdot\cap\Xi^0_i]\le C_\bullet\, E[\cdot]$. Let us denote

$$Z_i=\sum_{\tau_k\in\mathcal{D}:\mathbf{p}(\tau_k)\notin\boldsymbol{\Gamma}_i}[Jh(\tau_k)(|p(\tau_k)|^{-1}\wedge 1)]^2\Big(a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big),[(\mathcal{E}_0^{\tau_k})^*]^{-1}v\Big)^2_{\mathbb{R}^m}$$

and estimate $E[\varepsilon^{s,\mathbf{u}}Z_i]^{-\alpha-\delta}$, where $\varepsilon^{s,\mathbf{u}}$ denotes the operator adding the point $(s,\mathbf{u})$ to the configuration.

For $D\equiv[D(a,r)\wedge 1]$ ($D(a,r)$ is given in Definition 1.3), we have

$$\varepsilon^{s,\mathbf{u}}Z_i\ge C_\bullet\sum_{\tau_k\in\mathcal{D}:\mathbf{p}(\tau_k)\notin\boldsymbol{\Gamma}_i}[p(\tau_k)]^{2r}\mathbb{1}_{|p(\tau_k)|\le D}\mathbb{1}_{\tau_k\in[\beta,t-\beta]}, \tag{5.12}$$

here we used that $\mathcal{E}_0^{\cdot}$ is separated both from 0 and from $+\infty$ by some non-random constants.

Denote

$$A_i(\varkappa)\equiv\Big\{\{\tau_k\in\mathcal{D}\cap[\beta,t-\beta]:\mathbf{p}(\tau_k)\notin\boldsymbol{\Gamma}_i,|p(\tau_k)|\le D,|p(\tau_k)|>\varkappa\}=\emptyset\Big\},\quad \varkappa>0.$$

Due to the Chebyshev inequality, we have

$$P(\varepsilon^{s,\mathbf{u}}Z_i<C_\bullet\,\varkappa^{2r})\le$$

$$\le C_\bullet\,E\exp\left\{-\varkappa^{-2r}\sum_{\tau_k\in\mathcal{D}\cap[\beta,t-\beta]:\mathbf{p}(\tau_k)\notin\boldsymbol{\Gamma}_i}[p(\tau_k)]^{2r}\mathbb{1}_{|p(\tau_k)|\le D}\mathbb{1}_{|p(\tau_k)|\le\varkappa}\right\}\mathbb{1}_{A_i(\varkappa)}=$$

$$= C_\bullet E \prod_{\tau_k\in\mathcal{D}\cap[\beta,t-\beta]:\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i,|p(\tau_k)|\le D} \Psi(\varkappa,\tau_k), \tag{5.13}$$

where

$$\Psi(\varkappa,\tau_k) = \begin{cases} \exp\left\{-\varkappa^{-2r}[p(\tau_k)]^{2r}\right\}, & |p(\tau_k)|\le\varkappa, \\ 0, & |p(\tau_k)|>\varkappa. \end{cases}$$

Denote

$$\phi(\varkappa) = E \prod_{\tau_k\in\mathcal{D}\cap[\beta,(t-\beta)]:\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i,|p(\tau_k)|\le D} \Psi(\varkappa,\tau_k),$$

$$\phi^n(\varkappa) = E \prod_{\tau_k\in\mathcal{D}^n\cap[\beta,(t-\beta)]:\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i,|p(\tau_k)|\le D} \Psi(\varkappa,\tau_k),$$

we have $\phi^n\to\phi, n\to+\infty$. We may assume that the (locally finite) set $\{\tau_k\}=\mathcal{D}^n$ is ordered in the natural monotonous way. Denote, by Π_i, the projection on the first coordinate of the measure $\mathbf{\Pi}_i(\cdot)=\mathbf{\Pi}(\cdot\backslash\mathbf{\Gamma}_i)$. For every k ($\tau_k\in\mathcal{D}^n$) the value of the jump $p(\tau_k)$ is independent of $\mathcal{F}_k\equiv\mathcal{F}_{\tau_k-}\vee\sigma(\tau_k)$, and the distribution of the jump is equal to $[\Pi_i(\{|u|\ge\frac{1}{n}\})]^{-1}\cdot\Pi_i(\cdot\cap\{|u|\ge\frac{1}{n}\})$. Take $\varkappa<D, n>\frac{1}{\varkappa}$ and denote $\gamma_i^n=\Pi_i(\{|u|\ge\frac{1}{n}\})$. Then

$$E[\Psi(\varkappa,\tau_k)\mathbf{I}_{\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i,|p(\tau_k)|\le D}|\mathcal{F}_k]\le[\gamma_i^n]^{-1}\int_{\frac{1}{n}\le|u|\le\varkappa}\exp\{-\varkappa^{-2r}u^{2r}\}\Pi_i(du)=$$

$$=1-[\gamma_i^n]^{-1}\left\{\Pi_i(|u|\ge\varkappa)+\int_{\frac{1}{n}\le|u|\le\varkappa}[1-\exp\{-\varkappa^{-2r}u^{2r}\}]\Pi_i(du)\right\}. \tag{5.14}$$

It follows from (5.14) that

$$\phi^n\le E\left[1-[\gamma_i^n]^{-1}\left\{\Pi_i(|u|\ge\varkappa)+\int_{\frac{1}{n}\le|u|\le\varkappa}[1-\exp\{-\varkappa^{-2r}u^{2r}\}]\Pi_i(du)\right\}\right]^{N(n,i,D,\beta)}, \tag{5.15}$$

where $N(n,i,D,\beta)=\#\{k|\tau_k\in[\beta,t-\beta],\frac{1}{n}\le|p(\tau_k)|\le D\}$ is the Poissonian random variable with its intensity equal to $\gamma(n,i,D,\beta)\equiv(t-2\beta)\Pi_i(\frac{1}{n}\le|u|\le D)$. We have $\frac{\gamma(n,i,D,\beta)}{\gamma_n}\to(t-2\beta)$, and thus (5.15) implies that

$$\phi(\varkappa)\le\lim\sup_{n\to+\infty}\phi^n(\varkappa)\le\exp\left\{-(t-2\beta)\left[\Pi_i(|u|\ge\varkappa)+\int_{|u|\le\varkappa}[1-\exp\{-\varkappa^{-2r}u^{2r}\}]\Pi_i(du)\right]\right\}. \tag{5.16}$$

It follows from the construction of the grid that

$$\Pi_i(\cdot)\ge\frac{B-1}{B}\Pi(\cdot), \tag{5.17}$$

because while one cell $\mathbf{\Gamma}_i$ is removed, the "row" with the number $n(i)$ still contains $K_{n(i)}-1$ "copies" of this cell. Then, using (5.16) and the elementary inequality $1-\exp(-x)\ge\frac{e-1}{e}x, x\in[0,1]$, we obtain that

$$\phi(\varkappa)\le\exp\left\{-(t-2\beta)\frac{e-1}{e}\frac{B-1}{B}\left[\Pi(|u|>\varkappa)+\varkappa^{-2r}\int_{|u|\le\varkappa}u^{2r}\Pi(du)\right]\right\}=$$

$$=\exp\left\{-(t-2\beta)\frac{e-1}{e}\frac{B-1}{B}\ln\Big[\frac{1}{\varkappa}\Big]\rho_{2r}(\varkappa)\right\}=\varkappa^{(t-2\beta)\frac{e-1}{e}\frac{B-1}{B}\rho_{2r}(\varkappa)},$$

and consequently, for $\kappa=\varkappa^{2r}$,

$$P(\varepsilon^{s,\mathbf{u}}Z_i<C_\bullet\,\kappa)\le C_\bullet\,\kappa^{\frac{t-2\beta}{2r}\frac{e-1}{e}\frac{B-1}{B}\rho_{2r}(\kappa^{\frac{1}{2r}})}. \tag{5.18}$$

Now we put $\delta=\frac{1}{2}[t\frac{e-1}{2er}\boldsymbol{\rho}_{2r}-\alpha]$ and choose β and B in such a way that $\frac{t-2\beta(B-1)}{2rB}\frac{e-1}{e}\boldsymbol{\rho}_{2r}>\alpha+\frac{4\delta}{3}$. Then (5.18) implies that

$$\lim_{\varkappa\to0+}\sup_{\gamma,i,s,\mathbf{u}}\varkappa^{-\alpha-\delta}P(\varepsilon^{s,\mathbf{u}}Z_i<\kappa)<+\infty,$$

that proves the needed statement. The lemma is proved.

Let i be fixed. We can write

$$EF'(f)\frac{g_i^2}{Z+c} = P(\Xi)\frac{\lambda_i}{1+\lambda_i}\mathsf{E}_i^1 F'(f)\frac{g_i^2}{Z+c},$$

since $g_i = 0$ on $\Omega\backslash\Xi_i^1$. Using (5.8), we write

$$\mathsf{E}_i^1 F'(f)\frac{g_i^2}{Z+c} = \frac{1}{t}\int_{\mathbf{\Gamma}_i}\mathsf{E}_i^0\left[\int_0^t F'(f_{\mathbf{u}}(s))\frac{g_{i,\mathbf{u}}^2(s)}{Z_{\mathbf{u}}(s)+c}ds\right]\mu_i(d\mathbf{u}) =$$

$$= \frac{1}{t}\int_{\mathbf{\Gamma}_i}\mathsf{E}_i^0\left[\int_0^t F'(f_{\mathbf{u}}(s))G_{i,\mathbf{u}}(s)Y_{i,\mathbf{u},c}(s)ds\right]\mu_i(d\mathbf{u}), \tag{5.19}$$

where the following notation is used: $f_{\mathbf{u}}(s) = \varepsilon_i^{s,\mathbf{u}}f, g_{i,\mathbf{u}}(s) = \varepsilon_i^{s,\mathbf{u}}g_i, G_{i,\mathbf{u}}(s) = [Jh_i(s)]^{-1}g_{i,\mathbf{u}}(s), Z_{\mathbf{u}}(s) = \varepsilon_i^{s,\mathbf{u}}Z, Y_{i,\mathbf{u},c}(s) = [Jh_i(s)]^2\cdot\frac{G_{i,\mathbf{u}}(s)}{Z_{\mathbf{u}}(s)+c}$.

We are going to write the integration-by-parts formula for the integral w.r.t. ds in (5.19). In order to do this, we need some notation and preliminary results.

Definition 5.1. The function $f:\mathbb{R}^+\mapsto\mathbb{R}$ is called to belong to the class ACPD (absolutely continuous + purely discontinuous) if $f\in BV_{loc}(\mathbb{R}^+)$ and there exists the function $g\in L_{1,loc}(\mathbb{R}^+)$ such that

$$f(r-)-f(0+) = \int_0^r g(s)\,ds + \sum_{s\in(0,r)}[f(s+)-f(s-)],\quad r\in\mathbb{R}^+.$$

The function g λ^1-a.s. coincides with the derivative of f. Therefore we denote $g = f' = \frac{\partial}{\partial s}f$.

If f belongs to ACPD and is continuous, then it is absolutely continuous. In this case, we say that it belongs to the class AC.

The following statement is quite standard, and therefore we just outline its proof.

Proposition 5.3. *Let $f_1,\dots,f_m$ belong to the class ACPD. Then, for every $F\in C^1(\mathbb{R}^m)$, the function $F(f_1,\dots,f_m)$ belongs to the same class with*

$$[F(f_1,\dots,f_m)]' = \sum_{k=1}^m F_k'(f_1,\dots,f_m)f_k',$$

$$[F(f_1,\dots,f_m)](s+) - [F(f_1,\dots,f_m)](s-) = [F(f_1(s+),\dots,f_m(s+))] - [F(f_1(s-),\dots,f_m(s-))]$$

(the first equality should be understood in the λ^1-a.s. sense).

Sketch of the proof. The statement of the proposition is trivial when $f_1,\dots,f_m$ have only finite family $\{s_1<\dots<s_m\}$ of the points of discontinuity, and belong to the class C^1 on every interval $[s_k,s_{k+1}], k = 1,\dots,m-1$. If the functions $f_1,\dots,f_m$ belong to the class AC on every interval $[s_k,s_{k+1}]$, then one can prove the needed statement for them, approximating them, together with their derivatives, in L_1 sense on these intervals by smooth functions, and then passing to the limit. In the general case, one should first approximate every function f_j by the functions $f_j^\varepsilon, \varepsilon>0$, defined by the relations

$$f_j^\varepsilon(r-)-f_j^\varepsilon(0+) = \int_0^r f_j'(s)\,ds + \sum_{s\in(0,r)}[f(s+)-f(s-)]\mathbb{1}_{|f(s+)-f(s-)|>\varepsilon},$$

and then again pass to the limit as $\varepsilon\to 0+$.

Proposition 5.4. *There exist the modifications of the processes* $X(\cdot), \mathcal{E}_0^{\cdot}$ *such that, for any* $\mathbf{u} \in \mathbf{\Gamma}_i$,

1) for every $r \in [0,t]$, *the function* $s \mapsto \varepsilon_i^{s,\mathbf{u}} X(r)$ *belongs to AC with its derivative equal to*

$$\frac{\partial}{\partial s}\varepsilon_i^{s,\mathbf{u}} X(r) = (\varepsilon_i^{s,\mathbf{u}}\mathcal{E}_s^r)\Big[a\Big(X(s-)+u\Big) - a\Big(X(s-)\Big)\Big]\mathbf{1}_{[0,r]}(s), \quad s \in [0,t];$$

2) for every $r \in [0,t]$, *the function* $s \mapsto \varepsilon_i^{s,\mathbf{u}}\mathcal{E}_0^r$ *belongs to AC with*

$$\frac{\partial}{\partial s}\varepsilon_i^{s,\mathbf{u}}\mathcal{E}_0^r = (\varepsilon_i^{s,\mathbf{u}}\mathcal{E}_0^r)\Big[a\Big(X(s-)+u\Big) - a\Big(X(s-)\Big)\Big]\int_s^r a''(\varepsilon_i^{s,\mathbf{u}}X(z))\,dz \cdot \mathbf{1}_{[0,r]}(s), \quad s \in [0,t];$$

3) the function $s \mapsto \mathcal{E}_0^s$ *belongs to AC with* $\frac{\partial}{\partial s}\mathcal{E}_0^s = a'(X(s-))\mathcal{E}_0^s$;

4) the function $s \mapsto X(s-)$ *belongs to ACPD with* $\frac{\partial}{\partial s}X(s-) = \tilde{a}(X(s-)), \quad \tilde{a}(x) \equiv a(x) - \int_{|u|\le 1} u\,\Pi(du)$. *The set of jumps of this function coincides with* $\{s_j | \theta_j = 1\}$, *and the value of the jump at the point* s_j *is equal to* u_j.

Proof. Statements 3),4) follow straightforwardly from the construction of $X(\cdot), \mathcal{E}_0^{\cdot}$. Statement 1) is just the statement of Theorem 4.1 reformulated to the other form. Statement 2) follows from the considerations completely analogous to those given in the proof of Theorem 4.1. The proposition is proved.

As a corollary, we obtain the following statement.

Proposition 5.5. *There exist the modifications of the functions* f, g_i *such that, everywhere on* Ω_i^0 *for every* $\mathbf{u} \in \mathbf{\Gamma}_i$, *the function* $Y_{i,\mathbf{u},c}(\cdot)$ *belongs to the class ACPD, and the following integration-by-parts formula holds:*

$$\int_0^t F'(f_{\mathbf{u}}(s))G_{i,\mathbf{u}}(s)Y_{i,\mathbf{u},c}(s)ds = -\int_0^t F(f_{\mathbf{u}}(s))[Y_{i,\mathbf{u},c}]'(s)\,ds - \sum_{s\in[0,t]} F(f_{\mathbf{u}}(s))\big[Y_{i,\mathbf{u},c}(s+) - Y_{i,\mathbf{u},c}(s-)\big]. \tag{5.20}$$

Proof. It follows from Proposition 5.8 that $[f_{\mathbf{u}}]' = G_{i,\mathbf{u}}$ belongs to ACPD with

(5.21)

$$|G_{i,\mathbf{u}}(s)| \le C_2(a)|u|, \quad |[G_{i,\mathbf{u}}]'(s)| \le C_2(a)|u|(1+|X(s-)|), \quad |G_{i,\mathbf{u}}(s_j+) - G_{i,\mathbf{u}}(s_j-)| \le C_2(a)|u||u_j|, \quad j \ne i,$$

where the constant $C_2(a)$ depends only on $\|a'\|_\infty, \|a''\|_\infty$. Analogously, for $j \ne i$, the function

$$s \mapsto G_{i,j,\mathbf{u}}(s) = (\varepsilon_i^{s,\mathbf{u}}\mathcal{E}_{s_j}^t)\Big[a\Big(\varepsilon_i^{s,\mathbf{u}}X(s_j-) + u_j\Big) - a\Big(\varepsilon_i^{s,\mathbf{u}}X(s_j-)\Big)\Big]\mathbf{1}_{\{\theta_j=1\}}$$

belongs to AC with

$$|G_{i,j,\mathbf{u}}(s)| \le C_2(a)|u_j|, \quad |[G_{i,j,\mathbf{u}}]'(s)| \le C_2(a)|u||u_j|. \tag{5.22}$$

Then the function $\sum_{j\ne i}[Jh_j(s_j)]^2 G_{i,j,\mathbf{u}}^2(\cdot)$ belongs to AC with its derivative dominated by $|u|(C_2(a)\cdot\|Jh\|_\infty)^2\xi$, where

$$\xi = 2\sum_{j\in\mathbb{N}}(u_j^2 \wedge 1)\mathbf{1}_{\{\theta_j=1\}} \in \bigcap_{p>1} L_p(\Omega, M). \tag{5.23}$$

Therefore the function

$$Z_{\mathbf{u}}(\cdot) = G_{i,\mathbf{u}}^2(\cdot)[Jh_i(\cdot)]^2 + \sum_{j\ne i}[Jh_j(s_j)]^2 G_{i,j,\mathbf{u}}^2(\cdot)$$

belongs to the class ACPD. At last, $Z_{\mathbf{u}}(s) + c \ge c > 0$, and, applying Proposition 5.3 with $F \in C^1(\mathbb{R}^2)$ such that $F(x,y) = \frac{x}{y}$ for $x \in \mathbb{R}, y > c$, we obtain that $Y_{i,\mathbf{u},c}$ belongs to ACPD. Applying once again Proposition 5.3, we obtain (5.20) (we use here that $Jh_i(0) = Jh_i(t) = 0$, and thus $Y_{i,\mathbf{u},c}(0+) = Y_{i,\mathbf{u},c}(t-) = 0$). Proposition is proved.

Estimates (5.21),(5.22) straightforwardly imply the following estimates for $[Y_{i,\mathbf{u},c}]'$ and $\big[Y_{i,\mathbf{u},c}(s) - Y_{i,\mathbf{u},c}(s-)\big]$ that do not involve c.

Proposition 5.6. *1) For every $s\in[0,t]$,*

$$|[Y_{i,\mathbf{u},c}]'(s)|\le 2(|u|\wedge 1)(C_2(a)\cdot\|Jh\|_\infty)^2(\xi+1+|X(s-)|)[\Sigma_i^{\mathbf{u}}(s)]^{-2}.$$

2) For every $j\neq i$,

$$\left|Y_{i,\mathbf{u},c}(s_j)-Y_{i,\mathbf{u},c}(s_j-)\right|\le C_2(a)(|u|\wedge 1)(|u_j|\wedge 1)[\Sigma_i^{\mathbf{u}}(s_j)]^{-1}\mathbb{1}_{\{\theta_j=1\}}.$$

The constant $C_2(a)$ depends only on $\|a'\|_\infty,\|a''\|_\infty$.

Now we can write down the integration-by-parts formula for the functionals of $f=X(t)$ on (Ξ^γ,P^γ). Denote, by E^γ, the expectation w.r.t. P^γ and put $Y_{i,\mathbf{u}}\equiv Y_{i,\mathbf{u},0}$.

Lemma 5.3. *Let $a\in\mathbf{K}_r$ and $\frac{t}{2r}\frac{e-1}{e}\boldsymbol{\rho}_{2r}>2$ for some $r\in\mathbb{N}$. Suppose that the constants β,B in the construction of the grids $\mathcal{G}^\gamma$ are given by Lemma 5.2 with $\alpha=2,k=2$. Then*
(5.24)

$$E^\gamma F'(f)=-\frac{1}{t}\sum_{i\in\mathbb{N}}\frac{\lambda_i}{\lambda_i+1}\int_{\mathbf{\Gamma}_i}\mathsf{E}_i^0\left[\int_0^1 F(f_{\mathbf{u}}(s))Y'_{i,\mathbf{u}}(s)\,ds+\sum_{j\neq i}F(f_{\mathbf{u}}(s_j))[Y_{i,\mathbf{u}}(s_j+)-Y_{i,\mathbf{u}}(s_j-)]\right]\mu_i(d\mathbf{u})$$

for every $F\in C_b^1(\mathbb{R})$, and

$$\sup_\gamma\sum_{i\in\mathbb{N}}\frac{\lambda_i}{\lambda_i+1}\int_{\mathbf{\Gamma}_i}\mathsf{E}_i^0\left[\int_0^1|Y'_{i,\mathbf{u}}(s)|\,ds+\sum_{j\neq i}|Y_{i,\mathbf{u}}(s_j+)-Y_{i,\mathbf{u}}(s_j-)|\right]\mu_i(d\mathbf{u})<+\infty. \tag{5.25}$$

Remark. Two terms on the right-hand side of (5.24) can be naturally interpreted as the integrals of $F(f)$ w.r.t. some signed measures. Estimate (5.25) shows that these measures have finite total variation. The essential point here is that the second term in the integral w.r.t. the measure that is, in fact, singular w.r.t. the initial probability. This motivates us to call (5.24) the *singular type* integration-by-parts formula.

Proof. We have $\sup_{s\in[0,t]}E|\xi+1+X(s-)|^p<+\infty$ for every $p<+\infty$, thus statement 1) of Proposition 5.6 and Lemma 5.2 provide that

$$\int_{\mathbf{\Gamma}_i}\mathsf{E}_i^0\int_0^1|Y'_{i,\mathbf{u}}(s)|\,ds\mu_i(d\mathbf{u})\le C_\bullet\,(\varepsilon_{n(i)}\wedge 1),\quad i\in\mathbb{N},c>0.$$

Next, we use statement 2) of Proposition 5.4 and Proposition 5.2 to write

$$\int_{\mathbf{\Gamma}_i}\mathsf{E}_i^0\sum_{j\neq i}\left|Y_{i,\mathbf{u},c}(s_j)-Y_{i,\mathbf{u},c}(s_j-)\right|ds\mu_i(d\mathbf{u})\le$$

$$\le\frac{C_2(a)}{t}(\varepsilon_{n(i)}\wedge 1)\int_{\mathbf{\Gamma}_i}\sum_{j\neq i}\int_0^t\int_{\mathbf{\Gamma}_j}\mathsf{E}_{i,j}^0\frac{\lambda_j}{1+\lambda_j}(\varepsilon_{n(j)}\wedge 1)[\Sigma_{i,j}^{\mathbf{u},\tilde{\mathbf{u}}}(s,\tilde{s})]^{-1}\mu_j(d\tilde{\mathbf{u}})d\tilde{s}\mu_i(d\mathbf{u})\le$$

$$\le C_2(a)(\varepsilon_{n(i)}\wedge 1)\left[\sum_j\lambda_j(\varepsilon_{n(j)}\wedge 1)\right]\left[\mathsf{E}_{i,j}^0\sup_{i,j,\mathbf{u},\tilde{\mathbf{u}},s,\tilde{s}}[\Sigma_{i,j}^{\mathbf{u},\tilde{\mathbf{u}}}(s,\tilde{s})]^{-1}\right]\le C_\bullet\,(\varepsilon_{n(i)}\wedge 1),\quad i\in\mathbb{N},c>0$$

(see (5.10 for the notation $Z_{i,j}^{\mathbf{u},\tilde{b}u}$). In the last inequality, we used Lemma 5.2 and the fact that, due to condition (1.4),

$$\sum_j\lambda_j(\varepsilon_{n(j)}\wedge 1)\le 2\int_{\mathbb{R}}(|u|\wedge 1)\Pi(du)<+\infty.$$

Once again, we use $\sum_i\lambda_i(\varepsilon_{n(i)}\wedge 1)<+\infty$ and deduce (5.24) and (5.25). The lemma is proved.

Remark. The explicit estimates given above show that there exists a constant $\mathsf{C}_1<+\infty$ such that, *for every* grid $\mathcal{G}^\gamma$ constructed in the way given above *for any* $\gamma>0$, the expression on the left-hand side of (5.25) is dominated by C_1.

The last thing we need to complete the proof of Theorem 1.3 is to iterate (5.24) in order to provide an estimate for $EF^{(n)}(f)$ in the terms of $\sup_x |F(x)|$ ($F^{(n)}$ denotes the n-th derivative of F). The essential point here is that the measure M_i^0 is also the product measure and possesses the constructions given before for the measure M.

Let us rewrite (5.24) to the form that is convenient to the further iterative procedure. For a given n, we denote, by $\Theta(n)$, the family of all partitions $\theta=(\theta_1,\dots,\theta_r)$ of the set $\{1,\dots,n\}$ into non-overlapping parts (for instance, $\Lambda(2)$ contains two partitions $(\{1\},\{2\})$ and $(\{1,2\})$). Denote also, by $\mathbb{N}_d^n$, the set of all vectors $i_1,\dots,i_n$ with all coordinates not equal to one another. For a given $\bar{i}\equiv(i_1,\dots,i_n)\in\mathbb{N}_d^n$, $\bar{u}\equiv(\mathbf{u}_1,\dots,\mathbf{u}_n)$, $\bar{s}\equiv(s^1,\dots,s^r)$, and a partition $\theta=(\theta_1=\{\theta_1^1,\dots,\theta_1^{l_1}\},\dots,\theta_r=\{\theta_r^1,\dots,\theta_r^{l_r}\})\in\Lambda(n)$, we denote

$$\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}=[\varepsilon_{i_{\theta_1^1}}^{s_1,\mathbf{u}_1}\circ\varepsilon_{i_{\theta_1^2}}^{s_1,\mathbf{u}_2}\circ\dots\circ\varepsilon_{i_{\theta_1^{l_1}}}^{s_1,\mathbf{u}_{l_1}}]\circ[\varepsilon_{i_{\theta_2^1}}^{s_2,\mathbf{u}_{l_1+1}}\circ\dots\circ\varepsilon_{i_{\theta_2^{l_2}}}^{s_2,\mathbf{u}_{l_1+l_2}}]\circ\dots\circ[\varepsilon_{i_{\theta_r^1}}^{s_r,\mathbf{u}_{n-l_r+1}}\circ\dots\varepsilon_{i_{\theta_r^{l_r}}}^{s_r,\mathbf{u}_n}].$$

Now, using the statement analogous to the one of Proposition 5.2, applied to E_i^0 instead of E, we can write (5.24) in the form

$$E^\gamma F'(f)=\sum_{\theta\in\Theta(2)}\sum_{\bar{i}\in\mathbb{N}_d^2}\int_{\mathbf{\Gamma}_{i_1}\times\mathbf{\Gamma}_{i_2}}\int_{[0,t]^{r(\theta)}}\mathsf{E}_{i_1,i_2}^0F(\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}f)\,Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})\,d\bar{s}\,[\mu_{i_1}\times\mu_{i_2}](d\bar{\mathbf{u}}),\tag{5.26}$$

where $r(\theta)$ is the number of the components in the partition θ, and the functions $Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}$ are either a derivative or a jump of the function $Y_{i,\mathbf{u}}$ (in the notation of (5.24)) multiplied by $-\frac{\lambda_{i_1}\lambda_{i_2}}{t^2(\lambda_{i_1}+1)(\lambda_{i_2}+1)}$ or $-\frac{\lambda_{i_1}\lambda_{i_2}}{t(\lambda_{i_1}+1)(\lambda_{i_2}+1)}$, respectively.

Take $F\in C_b^2(\mathbb{R})$ and apply (5.26) to $\tilde{F}=F'$. Then the terms of the type $\mathsf{E}_{i_1,i_2}^0F'(\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}f)\,Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})$ occur on the right-hand side of (5.26). For every such a term, we write

$$\mathsf{E}_{i_1,i_2}^0F'(\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}f)\,Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})=\sum_{i\neq i_1,i_2}\mathsf{E}_{i_1,i_2}^0F'(\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}f)\,Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})\cdot\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}\left[\frac{g_i^2}{\Sigma_{i_1,i_2}}\right]=$$

$$=\sum_{i\neq i_1,i_2}\frac{\lambda_i}{t(\lambda_i+1)}\int_{\mathbf{\Gamma}_i}\int_0^t\varepsilon_i^{s,\mathbf{u}}[F'(\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}f)\,Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})]\cdot\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}\left[\frac{g_i^2}{\Sigma_{i_1,i_2}}\right]\,ds\mu_i(d\mathbf{u}),\tag{5.27}$$

where $\Sigma_{i_1,i_2}=\sum_{i\neq i_1,i_2}g_i^2$. From Proposition 5.4, we get that the function $s\mapsto\varepsilon_i^{s,\mathbf{u}}\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}f$ belongs to AC with

$$\frac{\partial}{\partial s}\varepsilon_i^{s,\mathbf{u}}\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}f=[Jh_i(s)]^{-1}\varepsilon_i^{s,\mathbf{u}}\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}g_i.$$

The function $s\mapsto[Jh_i(s)]\varepsilon_i^{s,\mathbf{u}}\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}\left[\frac{g_i}{\Sigma_{i_1,i_2}}\right]$ belongs to ACPD with its derivative and jumps satisfying the estimates analogous to those given in Proposition 5.6, but with $\Sigma_i^{\mathbf{u}}$ replaced by $\varepsilon_i^{s,\mathbf{u}}\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}\Sigma_{i,i_1,i_2,i}$, where $\Sigma_{i,i_1,i_2}=\sum_{j\neq i,i_1,i_2}g_j^2$. At last, using Proposition 5.4 and the explicit form of $Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})$, one can verify that the function $s\mapsto\varepsilon_i^{s,\mathbf{u}}Y_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})$ also belongs to ACPD with its derivative and jumps dominated by

$$C_\bullet\,\xi\cdot\lambda_i\lambda_{i_1}\lambda_{i_2}(\varepsilon_{n(i)}\wedge1)(\varepsilon_{n(i_1)}\wedge1)(\varepsilon_{n(i_2)}\wedge1)[\Sigma_{i,i_1,i_2}]^{-3},$$

where the constant $C_\bullet$ depends only on the coefficient a, and the variable ξ belongs to $\cap_pL_p$. This means that, under an appropriate moment condition imposed on $[\Sigma_{i,i_1,i_2}]^{-3}$, we can write the integration-by-parts formula on the right-hand side of (5.27) and obtain the analog of (5.26) with $E^\gamma F''(f)$ on the left-hand side. Let us formulate this statement for the derivative of an arbitrary order. For a given $\bar{i}\in\mathbb{N}_d^n$, we denote $\mathbf{\Gamma}_{\bar{i}}=\mathbf{\Gamma}_{i_1}\times\dots\times\mathbf{\Gamma}_{i_n}$, $\mu_{\bar{i}}=\mu_{i_1}\times\dots\times\mu_{i_n}$, $\mathsf{E}_{\bar{i}}^0=\mathsf{E}_{i_1,\dots,i_n}$, $\Sigma_{\bar{i}}=\sum_{i\not\in\bar{i}}g_i^2$.

Lemma 5.4. *Let $n\in\mathbb{N}$ be fixed, $a\in\mathbf{K}_r$ and $\frac{t}{2r}\frac{e-1}{e}\boldsymbol{\rho}_{2r}>2n$ for some $r\in\mathbb{N}$. Suppose that the constants β,B in the construction of the grids $\mathcal{G}^\gamma$ are given by Lemma 5.2 with $\alpha=2n,k=n$.*

Then there exists a set of the functions $\{Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}:[0,t]^{r(\theta)}\to\mathbb{R},\bar{i}\in\mathbb{N}^{2n}_d,\theta\in\Theta(2n),\bar{\mathbf{u}}\in[\mathbb{R}^2]^{2n}\}$ *such that*

$$E^\gamma F^{(n)}(f)=\sum_{\theta\in\Theta(2n)}\sum_{\bar{i}\in\mathbb{N}^{2n}_d}\int_{\mathbf{\Gamma}_{\bar{i}}}\int_{[0,t]^{r(\theta)}}\mathsf{E}^0_{\bar{i}}F(\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}f)\,Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}(\bar{s})\,d\bar{s}\,\mu_{\bar{i}}(d\bar{\mathbf{u}}),\tag{5.28}$$

and

(5.29)

$$\int_{\mathbf{\Gamma}_{\bar{i}}}\int_{[0,t]^{r(\theta)}}\mathsf{E}^0_{\bar{i}}|Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}(\bar{s})|\,d\bar{s}\,\mu_{\bar{i}}(d\bar{\mathbf{u}})\le C(n,\delta)\lambda_{i_1}\dots\lambda_{i_{2n}}\varepsilon_{n(i_1)}\dots\varepsilon_{n(i_{2n})}(1+\varepsilon_{n(i_1)})^{M(n,\delta)}\dots(1+\varepsilon_{n(i_{2n})})^{M(n,\delta)},$$

$\bar{i}\in\mathbb{N}^{2n}_d,\theta\in\Theta(2n)$, *where* $C(n,\delta),M(n,\delta)$ *are some constants depending only on* n *and the number* δ *given by Lemma 5.2.*

Proof. The iterative procedure described before shows how one can deduce formula (5.28) for a given n from the same formula for $n-1$: one should take one term in (5.28) and write down the formula analogous to (5.27) for it. This explains how the coefficients $Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}$ of the order n (i.e., with $\bar{i}\in\mathbb{N}^{2n}_d$) are constructed: one should take all $\bar{i}\in\mathbb{N}^{2n-2}_d$, $i\notin\bar{i}$ and calculate the derivative and the jump part of the function $s\mapsto\frac{\lambda_i[Jh_i(s)]^2}{t(\lambda_i+1)}\varepsilon^{s,\mathbf{u}}_i[Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}\frac{g_i}{\sum_{j\notin\bar{i}}g_j^2}]$. All such functions are exactly the new coefficients $Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}$. Such a description of the family $\{Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}\}$ allows one to rewrite it to the form $\{Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}(\bar{s})=H^{\bar{\mathbf{u}}}_{\bar{i},\theta}(\bar{s})[\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}\Sigma_{\bar{i}}]^{-2n}\}$, where the functions $\{H^{\bar{\mathbf{u}}}_{\bar{i},\theta}\}$ are defined iteratively. The power $2n$ here appears, since the power of the denominator increases by 1 twice on one step of the induction: the first time when the term $\frac{g_i}{\sum_{j\notin\bar{i}}g_j^2}$ is added, and the second one when either a derivative or the jump part is calculated.

Using the explicit expressions for the derivatives and jumps of the processes $X(\cdot),\mathcal{E}^{\cdot}_0$ (which the functions $\{g_i\}$, and thus the functions $\{H^{\bar{\mathbf{u}}}_{\bar{i},\theta}\}$, are expressed through) one can deduce by induction on n that, for every index sets $\bar{i}\in\mathbb{N}^n_d,\bar{j}$ with $\bar{l}=\bar{i}\cup\bar{j}=l_1,\dots,l_N$, for every ordered sets $p=(p_1,\dots,p_k)\subset\bar{l}^k,o=(o_1,\dots,o_k)\in\{0,1\}^k$,

$$|\partial^{o_1}_{p_1}\dots\partial^{o_k}_{p_k}\varepsilon^{\bar{s}^1,\bar{\mathbf{u}}^1}_{\bar{j}\setminus\bar{i}}H^{\bar{\mathbf{u}}}_{\bar{i},\theta}(\bar{s})|\le C_\bullet\,\varepsilon^{\bar{s}^1,\bar{\mathbf{u}}^1}_{\bar{j}\setminus\bar{i}}\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}[1+\max_{s\le t}|X(s)|]^{M(N,k)}\lambda_{l_1}\dots\lambda_{l_N}(\varepsilon_{n(l_1)}\wedge 1)\dots(\varepsilon_{n(l_N)}\wedge 1),\tag{5.30}$$

$\bar{i}\in\mathbb{N}^{2n}_d,\theta\in\Theta(2n),\bar{s}\in[0,t]^{r(\theta)},\bar{u}\in\mathbf{\Gamma}_{\bar{i}},\bar{s}^1\in[0,t]^{N-n},\bar{u}\in\mathbf{\Gamma}_{\bar{j}\setminus\bar{i}}$, where ∂^0_p denotes the derivative w.r.t. the variable with the number p, ∂^1_p denotes the jump w.r.t. the same variable, $\varepsilon^{\bar{s}^1,\bar{\mathbf{u}}^1}_{\bar{j}\setminus\bar{i}}\equiv\varepsilon^{\bar{s}^1,\bar{\mathbf{u}}^1}_{\bar{j}\setminus\bar{i},\theta_*}$ with $\theta_*=(\{1\},\dots,\{N-n\})$. We do not need estimate (5.30) in its full generality, we only need the partial case $\bar{j}=\bar{i},k=0$. In this case, we have the estimate

$$|H^{\bar{\mathbf{u}}}_{\bar{i},\theta}(\bar{s})|\le\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}[1+\max_{s\le t}|X(s)|]^{M(n)}\lambda_{i_1}\dots\lambda_{i_{2n}}(\varepsilon_{n(i_1)}\wedge 1)\dots(\varepsilon_{n(i_{2n})}\wedge 1),\tag{5.31}$$

$\bar{i}\in\mathbb{N}^{2n}_d,\theta\in\Theta(2n),\bar{s}\in[0,t]^{r(\theta)},\bar{u}\in\mathbf{\Gamma}_{\bar{i}}$, where $M(n)$ is some constant. Note that estimate (5.31) is not well designed to be proved by induction on n, while (5.30) is; this was the only reason for us to write firstly estimate (5.30). Now, using Lemma 5.2, we obtain

$$\int_{\mathbf{\Gamma}_{\bar{i}}}\int_{[0,t]^{r(\theta)}}\mathsf{E}^0_{\bar{i}}|Y^{\bar{\mathbf{u}}}_{\bar{i},\theta}(\bar{s})|\,d\bar{s}\,\mu_{\bar{i}}(d\bar{\mathbf{u}})\le\left[\int_{\mathbf{\Gamma}_{\bar{i}}}\int_{[0,t]^{r(\theta)}}\mathsf{E}^0_{\bar{i}}\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}[1+\max_{s\le t}|X(s)|]^{\frac{M(n)(2n+\delta)}{\delta}}d\bar{s}\mu_{\bar{i}}(d\mathbf{u})\right]^{\frac{\delta}{2n+\delta}}\times$$

$$\times\lambda_{i_1}\dots\lambda_{i_{2n}}(\varepsilon_{n(i_1)}\wedge 1)\dots(\varepsilon_{n(i_{2n})}\wedge 1).$$

Since ∇a is bounded and $\int_{\{|u|>1\}}|u|^p\Pi(du)<+\infty$ for every p, there exists such a constant $\tilde{C}(n)$ that

$$\mathsf{E}^0_{\bar{i}}\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}[1+\max_{s\le t}|X(s)|]^{\frac{M(n)(2n+\delta)}{\delta}}\le\left[\tilde{C}(n)(1+\|u_1\|^{M(n)})\dots(1+\|u_{2n}\|^{M(n)})\right]^{\frac{2n+\delta}{\delta}}.$$

This provides (5.29). The lemma is proved.

Now we can complete the proof of Theorem 1.3 in the case $m = 1$. We apply Lemma 5.4 for $n \le k+1$. Equality (5.28) and estimate (5.29) immediately imply that (5.7) holds true. Thus the needed statement holds true due to Lemma 5.1. The proof is complete.

Proof of Theorem 1.3: the case $m > 1$. All the technique, that is necessary for the proof of Theorem 1.3 in the general case, was already introduced in the proof of the case $m = 1$. Our aim now is to adapt this technique to the multidimensional situation.

Again, denote $f = X(t)1\!\!1_\Xi$, $g_i = D_{h_i}^{\mathbf{\Gamma}_i} f$, now f, g_i are the random vectors in $\mathbb{R}^m$. Considerations analogous to those given after estimate (5.9) show that $g = (g_i) \in L_2(\Omega, P, \mathbb{R}^m \otimes \ell_2)$ and $f \in W_2^1(\mathcal{G}^\gamma, \mathbb{R}^m)$ with $D^{\mathcal{G}} f = g$. We put

$$Z = \sum_i g_i \otimes g_i, \quad Z_{\bar{i}} = \sum_{i \notin \bar{i}} g_i \otimes g_i, \quad \bar{i} \in \mathbb{N}_d^n, n \ge 1,$$

Z is the Malliavin matrix for the vector f. We can write down the estimate analogous to (5.9) for $\|Z\|_{\mathbb{R}^{m^2}}$ and then prove (for instance, calculating the Fourier transform of the right-hand side and then estimating its derivatives of all the orders) that $\|Z\|_{\mathbb{R}^{m^2}} \in \cap_p L_p$.

We use the notation $\alpha \in \{1, \dots, m\}$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \{1, \dots, m\}^n$ for the indices and multiindices, $\partial_\alpha \equiv \frac{\partial}{\partial x_\alpha}, \partial_{\boldsymbol{\alpha}} \equiv \frac{\partial}{\partial x_{\alpha_1}} \dots \frac{\partial}{\partial x_{\alpha_n}}$. Let us write down the analogs of (4.21) and (4.25). First, we do this formally, without taking care of the terms involved in the corresponding integration-by-parts formula to belong to L_1. The necessary moment estimates will be given later on, in the second part of the proof.

Denote $Y_{i,\mathbf{u}}(s) = Jh_i(s)[\varepsilon_i^{s,\mathbf{u}} Z]^{-1} g_i, s \in [0,t], \mathbf{u} \in \mathbf{\Gamma}_i, i \in \mathbb{N}$, and let $Y_{i,\mathbf{u}}^\alpha$ denote the α-th component of the vector $Y_{i,\mathbf{u}}$. Using Proposition 5.3 and an appropriate analog of Proposition 5.5, one can obtain the following analog of the integration-by-parts formula (5.24):

(5.32)

$$E^\gamma[\partial_\alpha F](f) = -\frac{1}{t} \sum_{i \in \mathbb{N}} \frac{\lambda_i}{\lambda_i + 1} \int_{\Gamma_i} \mathsf{E}_i^0 \left[\int_0^1 F(\varepsilon_i^{s,\mathbf{u}} f)[Y_{i,\mathbf{u}}^\alpha]'(s)\, ds + \sum_{j \ne i} F(\varepsilon_i^{s,\mathbf{u}} f)[Y_{i,\mathbf{u}}^\alpha(s_j+) - Y_{i,\mathbf{u}}^\alpha(s_j-)] \right] \mu_i(d\mathbf{u}),$$

for every $F \in C_b^1(\mathbb{R}^m)$ and $\alpha \in \{1, \dots, m\}$. One can rewrite (5.32) to the form analogous to (5.26) and then iterate this formula in the way described before the formulation of Lemma 5.4. The inverse matrix Z^{-1} can be expressed in the form $[\det Z]^{-1} Q$, where the elements of the matrix Q (the *cofactor matrix* for Z) are certain polynomials of the elements of Z. At last, for every $\bar{i}_1 \subset \bar{i}_2$, $\det Z_{\bar{i}_2} \le \det Z_{\bar{i}_1}$. Summarizing all these considerations, we can formulate the following statement.

Proposition 5.7. *For every $F \in C_b^n(\mathbb{R}^m), n \ge 1, \boldsymbol{\alpha} \in \{1, \dots, m\}^n$,*

$$E^\gamma[\partial_{\boldsymbol{\alpha}} F](f) = \sum_{\theta \in \Theta(2n)} \sum_{\bar{i} \in \mathbb{N}_d^{2n}} \int_{\mathbf{\Gamma}_{\bar{i}}} \int_{[0,t]^{r(\theta)}} \mathsf{E}_{\bar{i}}^0 F(\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}} f)\, Y_{\bar{i},\theta}^{\bar{\mathbf{u}},\boldsymbol{\alpha}}(\bar{s})\, d\bar{s}\, \mu_{\bar{i}}(d\bar{\mathbf{u}}). \tag{5.33}$$

Here the family $\{Y_{\bar{i},\theta}^{\bar{\mathbf{u}},\boldsymbol{\alpha}}(\bar{s})\}$ possesses the point-wise representation $\{Y_{\bar{i},\theta}^{\bar{\mathbf{u}},\boldsymbol{\alpha}}(\bar{s}) = H_{\bar{i},\theta}^{\bar{\mathbf{u}},\boldsymbol{\alpha}}(\bar{s})[\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}} \det Z_{\bar{i}}]^{-2n}\}$ with the functions $\{H_{\bar{i},\theta}^{\bar{\mathbf{u}},\boldsymbol{\alpha}}\}$ estimated by

$$|H_{\bar{i},\theta}^{\bar{\mathbf{u}}}(\bar{s})| \le \varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}}[1 + \max_{s \le t} \|X(s)\|]^{M(n)} \lambda_{i_1} \dots \lambda_{i_{2n}} (\varepsilon_{n(i_1)} \wedge 1) \dots (\varepsilon_{n(i_{2n})} \wedge 1), \tag{5.34}$$

$\bar{i} \in \mathbb{N}_d^{2n}, \theta \in \Theta(2n), \bar{s} \in [0,t]^{r(\theta)}, \bar{u} \in \mathbf{\Gamma}_{\bar{i}}$, *where the constant $M(n)$ depends only on n.*

Equality (5.33) is now nothing more than the formal expression, since the variables $Y_{\bar{i},\theta}^{\bar{\mathbf{u}},\boldsymbol{\alpha}}$ may not belong to L_1. However, estimate (5.34) allows one to separate the case where this equality becomes meaningful and rigorous.

Corollary 5.1. *Suppose that the grids $\mathcal{G}^\gamma$ were constructed in such a way that, for some $n \in \mathbb{N}, \delta > 0$,*

$$\sup_\gamma \sup_{l\le 2n} \sup_{\bar{i}\in\mathbb{N}_d^l} \sup_{\bar{\mathbf{u}}\in\mathbf{\Gamma}_{\bar{i}}} \sup_{\bar{s}\in[0,t]^l} \mathsf{E}^0_{\bar{i}}[\varepsilon_{i_1}^{s_1,\mathbf{u}_1}\dots\varepsilon_{i_l}^{s_l,\mathbf{u}_l} \det Z_{\bar{i}}]^{-2n-\delta}. \tag{5.35}$$

Then (5.33) holds true with

$$\sup_\gamma \sum_{\theta\in\Theta(2n)} \sum_{\bar{i}\in\mathbb{N}_d^{2n}} \int_{\mathbf{\Gamma}_{\bar{i}}} \int_{[0,t]^{r(\theta)}} \mathsf{E}^0_{\bar{i}} |Y^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s})|\, d\bar{s}\, \mu_{\bar{i}}(d\bar{\mathbf{u}}) = \mathsf{C}_n < +\infty. \tag{5.36}$$

Thus, the only essential fact, that it is left to prove, is the following multidimensional analog of Lemma 5.2.

Lemma 5.5. *Let $a \in \mathbf{K}_r$ and $\frac{t}{2r}\frac{e-1}{e}\boldsymbol{\rho}_{2r} > (\alpha+4)m-4$ for some $r \in \mathbb{N}, \alpha \in [0,+\infty)$. Then, for every $k \in \mathbb{N}$, there exists $\delta > 0$ such that, under an appropriate choice of the constants B, β in the construction of the grids $\mathcal{G}^\gamma$,*

$$\sup_\gamma \sup_{l\le k} \sup_{i_1,\dots,i_l\in\mathbb{N}} \sup_{\mathbf{u}_1\in\mathbf{\Gamma}_{i_1},\dots,\mathbf{u}_l\in\mathbf{\Gamma}_{i_l}} \sup_{s_1,\dots,s_l\in[0,t]} \left[\mathsf{E}^0_{i_1,\dots,i_l}[\varepsilon_{i_1}^{s_1,\mathbf{u}_1}\dots\varepsilon_{i_l}^{s_l,\mathbf{u}_l} \det Z_{(i_1,\dots,i_l)}]^{-\alpha-\delta}\right] < +\infty.$$

Proof. We consider only the case $k = 1$, the general case is completely analogous. We have $g_i = \mathcal{E}_0^t q_i$, $q_i \equiv Jh_i(\tau_i)[\mathcal{E}_0^{\tau_i}]^{-1}\Big[a\Big(X(\tau_i-)+p(\tau_i)\Big) - a\Big(X(\tau_i-)\Big)\Big]\mathbf{1}_{\Xi_i^\gamma}$. Define

$$Q = \sum_i q_i \otimes q_i, \quad Q_i = \sum_{j\ne i} q_j \otimes q_j,$$

then $Z = \mathcal{E}_0^t \cdot Q \cdot [\mathcal{E}_0^t]^*, Z_i = \mathcal{E}_0^t \cdot Q_i \cdot [\mathcal{E}_0^t]^*$. Since ∇a is bounded, $|\det \mathcal{E}_0^t|$ is separated from 0 by some non-random constant (see Proposition 6.2 below for the explicit estimate). Thus, in order to prove the statement of the lemma for $k = 1$, it is enough to prove that

$$\sup_\gamma \sup_{i\in\mathbb{N}} \sup_{\mathbf{u}\in\mathbf{\Gamma}_i} \sup_{s\in[0,t]} \left[\mathsf{E}^0_i[\varepsilon_i^{s,\mathbf{u}} \det Q_i]^{-\alpha-\delta}\right] < +\infty. \tag{5.37}$$

The calculations given in the proof of Lemma 1 [18] provide that, in order to verify (5.37), it is enough to prove that

$$\sup_\gamma \sup_{i\in\mathbb{N}} \sup_{\mathbf{u}\in\mathbf{\Gamma}_i} \sup_{s\in[0,t]} \sup_{v:\|v\|=1} \left[\mathsf{E}^0_i[(\varepsilon_i^{s,\mathbf{u}} Q_i v, v)_{\mathbb{R}^m}]^{4-m(\alpha+4)-\delta}\right] < +\infty. \tag{5.38}$$

We do this analogously to the proof of Lemma 5.2. Let us return from the "censored" probability space (Ω_i^0, M_i^0) to the initial one (Ω, P) and estimate $E[(\varepsilon^{s,\mathbf{u}} Q_i v, v)]^{4-m(\alpha+4)-\delta}$, where $\varepsilon^{s,\mathbf{u}}$ denotes the operator adding the point $(s, \mathbf{u})$ to the configuration. We have

$$(Q_i v, v)_{\mathbb{R}^m} = \sum_{\tau_k\in\mathcal{D}:\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i} [Jh(\tau_k)]^2 \|(\mathcal{E}_0^{\tau_k})^*]^{-1}v\|^2 \left(a\Big(X(\tau_k-)+p(\tau_k)\Big) - a\Big(X(\tau_k-)\Big), \frac{[(\mathcal{E}_0^{\tau_k})^*]^{-1}v}{\|(\mathcal{E}_0^{\tau_k})^*]^{-1}v\|}\right)^2_{\mathbb{R}^m} [\|p(\tau_k)\|\wedge 1]^2.$$

Since ∇a is bounded,

$$\operatorname{essinf} \inf_{\|v\|=1} \|(\mathcal{E}_0^{\tau_k})^*]^{-1}v\| \ge C_\bullet > 0$$

for every k (see Proposition 6.2 below). Thus, we deduce that, for every $\varrho \in (0,1)$, the following inequality holds true for $D \equiv [D(a, r, \varrho) \wedge 1]$:

$$(Q_i v, v) \ge C_\bullet \sum_{\tau_k\in\mathcal{D}:\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i} \Big(p(\tau_k), w(\tau_k)\Big)^{2r}_{\mathbb{R}^m} \mathbf{1}_{p(\tau_k)\in V(w(\tau_k),\varrho)} \mathbf{1}_{|p(\tau_k)|\le D} \mathbf{1}_{\tau_k\in[\beta,t-\beta]}, \tag{5.39}$$

where we denoted $w(\tau) \equiv w(X(\tau-), \frac{[(\mathcal{E}_0^{\tau_k})^*]^{-1}v}{\|(\mathcal{E}_0^{\tau_k})^*]^{-1}v\|})$ (see Definition 1.3 for the notation $w(\cdot,\cdot)$). Denote

$A_i(\varkappa) \equiv \{\tau_k \in \mathcal{D}\cap[\beta, t-\beta] : \mathbf{p}(\tau_k) \notin \mathbf{\Gamma}_i, p(\tau_k) \in V(w(\tau_k), \varrho), |p(\tau_k)| \le D, |(p(\tau_k), \varepsilon^{s,\mathbf{u}} w(\tau_k))| > \varkappa\} = \emptyset\}, \varkappa > 0.$

Due to the Chebyshev inequality, we have

$$P((\varepsilon^{s,\mathbf{u}}Q_i v, v) < C_\bullet\, \varkappa^{2r}) \leq$$

$$\leq C_\bullet\, E \exp\left\{ -\varkappa^{-2r} \sum_{\tau_k \in \mathcal{D}\cap[\beta, t-\beta]: \mathbf{p}(\tau_k)\notin \mathbf{\Gamma}_i} \Big(p(\tau_k), \varepsilon^{s,\mathbf{u}} w(\tau_k)\Big)^{2r}_{\mathbb{R}^m} \mathbf{1}_{p(\tau_k)\in V(w(\tau_k),\varrho)} \mathbf{1}_{|p(\tau_k)|\leq D} \mathbf{1}_{|(p(\tau_k), \varepsilon^{s,\mathbf{u}} w(\tau_k))|\leq \varkappa} \right\} \times$$

$$\times \mathbf{1}_{A_i(\varkappa)} = C_\bullet\, E \prod_{\tau_k \in \mathcal{D}\cap[\beta, t-\beta]: \mathbf{p}(\tau_k)\notin \mathbf{\Gamma}_i, p(\tau_k)\in V(w(\tau_k),\varrho), |p(\tau_k)|\leq D} \Psi(\varkappa, \tau_k),$$

where

$$\Psi(\varkappa, \tau_k) = \begin{cases} \exp\left\{ -\varkappa^{-2r} \Big(p(\tau_k), \varepsilon^{s,\mathbf{u}} w(\tau_k)\Big)^{2r}_{\mathbb{R}^m} \right\}, & |(p(\tau_k), \varepsilon^{s,\mathbf{u}} w(\tau_k))| \leq \varkappa, \\ 0, & |(p(\tau_k), \varepsilon^{s,\mathbf{u}} w(\tau_k))| > \varkappa. \end{cases}$$

One has that $p(\tau_k)$ is independent of $\mathcal{F}_k \equiv \mathcal{F}_{\tau_k -} \vee \sigma(\tau_k)$, and $w(\tau_k)$ is $\mathcal{F}_k$-measurable. Thus, repeating the arguments given in the proof of Lemma 5.2, one can obtain analogously to (5.14 – 5.18) that

$$P((\varepsilon^{s,\mathbf{u}}Q_i v, v)_{\mathbb{R}^m} < C_\bullet\, \kappa) \leq C_\bullet\, \kappa^{\frac{t-2\beta}{2r} \frac{e-1}{e} \frac{B-1}{B} \rho_{2r}(\kappa^{\frac{1}{2r}}, \varrho)}, \quad \varkappa \in (0, D),$$

and, under an appropriate choice of ϱ, β, B,

$$\lim_{\varkappa\to 0+} \sup_{\gamma, i, s, \mathbf{u}, \|v\|=1} \varkappa^{4-m(\alpha+4)-\delta} P((\varepsilon^{s,\mathbf{u}}Q_i v, v)_{\mathbb{R}^m} < \varkappa) = 0$$

for $\delta = \frac{1}{2}\left[\frac{t}{2r}\frac{e-1}{e}\boldsymbol{\rho}_{2r} - (\alpha+4)m + 4\right]$. The lemma is proved.

Corollary 5.2. *Under condition of Theorem 1.3, the grids $\mathcal{G}^\gamma$ can be constructed in such a way that the integration-by-parts formula (5.33) together with the moment estimate (5.36) hold true for $n \leq k + m$.*

This corollary immediately implies that estimates (5.7) hold true for $n \leq m + k$. Now the statement of Theorem 1.3 follows from Lemma 5.1. The theorem is proved.

Let us make a conclusive remark. The first and second terms in the integration-by-parts formula (5.24) can be interpreted as the "volume integral" and "surface integral", respectively, since the measure in the second term is supported, in fact, by the countable union of the sets $I_{i,j} \equiv \{s_i = s_j\}, i, j \in \mathbb{N}$, and each of these sets can be interpreted as a "level set" (or "codimension 1 set"). This is the main reason for the calculus of variations, developed in this section, to be substantially different from the classical (Malliavin's) form of the stochastic calculus of variations, since, in the latter one, the new measure is absolutely continuous w.r.t. the initial one, i.e. in the integration-by-parts formula only the "volume integral" is present.

It should be mentioned that the differential structure in our case is not like the one for the manifold with a (smooth) boundary. The "surface measure" again admits the similar regular structure, and the integration-by-parts formula for such a measure generates the "codimension 1" and "codimension 2" terms, and so on. Thus one can informally say that the phase space of the Poisson random measure, considered with the differential structure generated by the time-stretching transformations, looks like the "infinite-dimensional complex". The crucial point in our construction is that, on every "side of codimension k" of such a complex, there still remains an infinite family of admissible directions.

6. Smoothness of the density of the invariant distribution

In this section, we consider the stationary process $\{X(s), s\in\mathbb{R}\}$ satisfying the equation

$$X(t)-X(s)=\int_s^t a(X(r))\,dr+U_t-U_s,\quad -\infty<s\le t<+\infty, \tag{6.1}$$

with the Lévy process U defined on $\mathbb{R}$ by the standard construction

$$U_t=\begin{cases}U_t^1, & t\ge 0\\ -U^2_{(-t)-}, & t<0\end{cases},$$

where U^1, U^2 are two independent copies of the Lévy process defined on $\mathbb{R}^+$. The coefficient a is supposed to satisfy the conditions formulated in subsection 1.3.

In order to prove the regularity of the distribution of $X(t)$ (i.e., the statement of Theorem 1.5), we need to modify slightly the constructions from the Sections 3 and 5. The reason is that now one cannot suppose the probability space $(\Omega,\mathcal{F},P)$ to satisfy the condition $\mathcal{F}=\sigma(U)$. Such a supposition is, in fact, the claim to (6.1) to possess a strong solution on $\mathbb{R}$ and is, in general, a non-trivial restriction. In order to avoid such a restriction, we make the following modifications of the constructions given above.

Denote $H=L_2(\mathbb{R})$. Let $H_0\subset L_\infty(\mathbb{R})$ be the set of functions with a bounded support. For $h\in H_0$ denote $Jh(\cdot)=\int_{-\infty}^{\cdot}h(s)\,ds, b(h)=\sup\{r|h(v)=0, v\le r\}$. For a fixed $h\in H_0$, we define the family $\{T_h^t, t\in\mathbb{R}\}$ of transformations of the axis $\mathbb{R}$ by putting $T_h^t x, x\in\mathbb{R}$ equal to the value at the point $s=t$ to the solution of the Cauchy problem (3.1).

For every $h\in H_0,\Gamma\in\Pi_{fin}$, the transformation T_h^Γ of the random measure ν associated with U is well defined. Since $T_h^t x\equiv x, x\le b(h)$, the transformation T_h^Γ does not change the values of ν on every subset of $(-\infty,b(h)]\times\mathbb{R}^m$. Equation (6.1) considered as the Cauchy problem with s fixed possesses the strong solution. Thus, one can define the transformation T_h^Γ of the process X in such a way that $T_h^\Gamma X(t)=X(t), t\le b(h)$,

$$T_h^\Gamma X(t)=X(b(h))+\int_{b(h)}^t a(T_h^\Gamma X(r))\,dr+T_h^\Gamma(U_t-U_{b(h)}),\quad t\ge b(h). \tag{6.2}$$

Like in the proof of Theorem 1.3, we enlarge the probability space and suppose that the random measure ν associated with the process U is the projection on the first m coordinates of the random measure $\boldsymbol{\nu}$ defined on $\mathbb{R}\times\mathbb{R}^{m+1}$, with its intensity measure being equal to $\lambda^1\times\mathbf{\Pi},\mathbf{\Pi}\equiv\Pi\times\left[\lambda^1|_{[0,1]}\right]$. One possible formal way to do this is to define $(\Omega,\mathcal{F},P)$ as the product of two probability spaces $(\Omega^1,\mathcal{F}^1,P^1)$, $(\Omega^2,\mathcal{F}^2,P^2)$, where $\mathcal{F}^1=\sigma(X)$, and $\Omega^2=[0,1]^\infty, P^2=\prod_{l\in\mathbb{N}}\left[\lambda^1|_{[0,1]}\right]$. We enumerate jumps of the process X in some measurable way and put

$$\mathbf{X}(t)=\begin{cases}(X(t),0), & X(t)=X(t-)\\ (X(t),\xi_{l(t)}), & X(t)\ne X(t-)\end{cases},$$

where $\{\xi_l\}$ is the sequence of coordinate functionals on Ω^2 (i.e., every ξ_l has uniform distribution on $[0,1]$), and $l(t)$ denotes the number of the jump that happens at the moment t. Then $\sigma(\mathbf{X})=\mathcal{F}$, and the random measure $\boldsymbol{\nu}$ and the corresponding point process $\mathbf{p}(\cdot)$ can be constructed from $\mathbf{X}$ in the obvious way. For every $h\in H_0,\mathbf{\Gamma}\in\mathbf{\Pi}_{fin}$, the transformation $T_h^{\mathbf{\Gamma}}$ of the process $\mathbf{X}$ is well defined (the first coordinate X is transformed accordingly to (6.2), and the transformation of the last coordinate $\xi_{l(t)}$ is defined by the condition $T_h^{\mathbf{\Gamma}}[l(t)]=l(T_{-h}t)$).

Further we suppose that $\mathcal{F}=\sigma(\boldsymbol{\nu})$. Under this condition, one can easily verify that an analog of Lemma 3.1 holds true, and $T_h^{\mathbf{\Gamma}}$ is, in fact, the admissible transformation of $(\Omega,\mathcal{F},P)$ (the explicit formula for p_h differs slightly from the one given in subsection 3.1). The notions of the stochastic and a.s. derivatives associated with such admissible transformations can be introduced, and then the statement of Theorem 4.1 holds true

for every given $h \in H_0$ with the trivial replacements: 0 should be replaced by $b(h)$ and x should be replaced by $X(b(h))$.

We introduce the notion of a differential grid in the same way with Definition 3.2, with $\mathbb{R}^+$ replaced by $\mathbb{R}$ and a_i claimed to belong to $\mathbb{R}$ (i.e., a_i should not be equal to $-\infty$) for every i. For every such a grid, the Sobolev classes associated with the grid are defined in the same way with Definition 3.3.

Now let us proceed with the proof of Theorem 1.5. Since X is a stationary process, it is enough to study the distribution of $X(t)$ at one fixed point t, say, $t = 0$. For every given $\gamma \in (0, \frac{1}{2})$, we construct the grid $\mathcal{G}^\gamma$ in the way analogous to one given at the beginning of subsection 5.2. We take the same sequence $\{\varepsilon_n\}$ and consider all sets of the type

$$[-N, -N+1) \times I_n \times [\frac{k-1}{K_{n,N}}, \frac{k}{K_{n,N}}), \quad k = 1, \dots, K_{n,N}, n \in \mathbb{Z}, N \in \mathbb{N}, \tag{6.3}$$

recall that $I_n = \{u \| \|u\| \in [\varepsilon_{n+1}, \varepsilon_n)\}$. We enumerate sets (6.3) by $i \in \mathbb{N}$ in an arbitrary way and denote, by $n(i), N(i)$ and $k(i)$, such numbers that the corresponding components in the set with the number i are equal to $[-N(i), -N(i)+1)$, $I_{n(i)}$, and $[\frac{k(i)-1}{K_{n(i),N(i)}}, \frac{k(i)}{K_{n(i),N(i)}})$. The numbers $K_{n,N}$ are defined for every given $B > 0, \gamma$ by

$$K_{n,N} = \left[\max\left(B, 2t\Pi(I_n), \frac{3}{\gamma} \cdot 2^{|n|-N-1} t^2 \Pi(I_n)\right)\right] + 2,$$

and therefore

1) $K_{n,N} \geq B$ (the constant B will be determined below);
2) $\frac{1}{K_{n,N}} \Pi(I_n) < \frac{1}{2}$;
3) $\frac{1}{K_{n,N}} \Pi^2(I_n) < \frac{2\gamma}{3} 2^{-|n|-N}$.

We define the grids $\mathcal{G}^\gamma$ by the equalities $[a_i^\gamma, b_i^\gamma) = [-N(i), -N(i)+1)$, $\mathbf{\Gamma}_i = I_{n(i)} \times [\frac{k(i)-1}{K_{n(i),N(i)}}, \frac{k(i)}{K_{n(i),N(i)}})$,

$$h_i^\gamma(s) = A^{-N(i)} (\varepsilon_{n(i)}^{-1} \wedge 1) h(s + N(i)), \quad s \in \mathbb{R},$$

where $A > 1$ will be determined later on, and $h \in C^\infty$ is some given function such that $Jh = 0$ outside $[0,1]$, $Jh > 0$ on $(0,1)$, and $Jh = 1$ on $[\frac{1}{3}, \frac{2}{3}]$.

The construction of the grids $\mathcal{G}^\gamma$ provides that the estimate analogous to (5.6) holds true. Next, for the function $f = X(0) \mathbb{1}_{\Xi^\gamma}$, the estimate analogous to (5.9) can be written, and one can prove that $f \in \cap_p W_p^1(\mathcal{G}^\gamma, \mathbb{R}^m)$ with

$$g_i \equiv D_{h_i}^{\mathbf{\Gamma}_i} f = Jh_i(\tau_i) \mathcal{E}_{\tau_i}^0 \left[a\Big(X(\tau_i -) + p(\tau_i)\Big) - a\Big(X(\tau_i -)\Big)\right] \mathbb{1}_{\Xi_i^\gamma}$$

(here and below, we use the notation from subsection 5.2). Repeating step-by-step the considerations given in subsection 5.2, we obtain the following analog of Proposition 5.7. Denote, by $S(\bar{i}, \theta)$ for $\bar{i} \in \mathbb{N}_d^{2n}, \theta \in \Theta(2n)$, the set of $(s_1, \dots, s_{2n}) \in (-\infty, 0)$ such that

1) $s_{i_k} \in [-N(i_k), -N(i_k)+1)$, $k = 1, \dots, 2n$;
2) $s_{i_k} = s_{i_j}$ for every k, j such that i_k, i_j belong to the same set w.r.t. the partition θ.

Denote by $\lambda_{\bar{i},\theta}$ the uniform distribution on $S(\bar{i}, \theta)$, i.e the surface measure on $S(\bar{i}, \theta)$ considered as a subset of $\mathbb{R}^{2n}$ with the Lebesgue measure λ.

Proposition 6.1. *For every $F \in C_b^n(\mathbb{R}^m), \boldsymbol{\alpha} \in \{1, \dots, m\}^n$*

$$E^\gamma[\partial_{\boldsymbol{\alpha}} F](f) = \sum_{\theta \in \Theta(2n)} \sum_{\bar{i} \in \mathbb{N}_d^{2n}} \int_{\mathbf{\Gamma}_{\bar{i}}} \int_{S(\bar{i},\theta)} \mathsf{E}_{\bar{i}}^0 F(\varepsilon_{\bar{i},\theta}^{\bar{s},\bar{\mathbf{u}}} f) \, Y_{\bar{i},\theta}^{\bar{\mathbf{u}},\boldsymbol{\alpha}}(\bar{s}) \, \lambda_{\bar{i},\theta}(d\bar{s}) \, \mu_{\bar{i}}(d\bar{\mathbf{u}}). \tag{6.4}$$

The family $\{Y^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s})\}$ *possesses the point-wise representation* $\{Y^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s}) = H^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s})[\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}\det Z_{\bar{i}}]^{-2n}\}$ *with the functions* $\{H^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}\}$ *estimated by*

(6.5)
$$|H^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s})| \le C_\bullet\, \varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}[1+\max_{s\in[\min(s_1,\dots,s_n),0]}|X(s)|]^{M(n)}\lambda_{i_1}\dots\lambda_{i_{2n}}(\varepsilon_{n(i_1)}\wedge 1)\dots(\varepsilon_{n(i_{2n})}\wedge 1)A^{-\max(N(i_1),\dots,N(i_{2n}))},$$

$\bar{i}\in\mathbb{N}^{2n}_d, \theta\in\Theta(2n), \bar{s}\in S(\bar{i},\theta), \bar{u}\in\mathbf{\Gamma}_{\bar{i}}$.

Remark. In the estimates for $|H^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s})|$ we dominate all terms of the type $Jh_i(\tau_i), [Jh_i]'(\tau_i),\cdots$, by some constant $C_\bullet$ except the terms of the same type with $N(i)=\max(N(i_1),\dots,N(i_{2n}))$. These terms are dominated by $C_\bullet A^{-\max(N(i_1),\dots,N(i_{2n}))}$, that provides the term $A^{-\max(N(i_1),\dots,N(i_{2n}))}$ in (6.5).

Let us repeat the cautions made after Proposition 5.7: equality (6.4) is just a formal one; in order to make it rigorous the proof that $Y^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s})$ are integrable w.r.t. $\mathsf{E}^0_{\bar{i}}$ is needed. Such a proof should contain two parts: the estimate of the moment of $\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}[1+\max_{s\in[\min(s_1,\dots,s_n),0]}|X(s)|]^{M(n)}$, and the estimate of the moment of $[\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}\det Z_{\bar{i}}]^{-2n}$.

The first part of the proof is more or less standard. The variable $\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}X(\min(s_1,\dots,s_n)-)$ is, in fact, equal to $X(\min(s_1,\dots,s_n)-)$, and thus its distribution w.r.t. $\mathsf{E}^0_{\bar{i}}$ is equal to the initial invariant distribution P^*. This distribution was supposed in the formulation of Theorem 1.5 to have all the moments. Moreover, the gradient ∇a is globally bounded, and thus we can deduce from the standard martingale inequalities and the Gronwall lemma that there exists a constant $C(a)\equiv\sup_x\|\nabla a(x)\|]$ such that, for every $p>1$,

$$\left[\mathsf{E}^0_{\bar{i}}\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}[1+\max_{s\in[\min(s_1,\dots,s_n),0]}|X(s)|]^p\right]^{\frac{1}{p}} \le C_\bullet\,(1+\|\varepsilon_{n(i_1)}\|)\dots(1+\|\varepsilon_{n(i_{2n})}\|)e^{-C(a)\min(s_1,\dots,s_n)} \tag{6.6}$$

with the constant $C_\bullet$ depending on p,n and the moments of P^*.

The second part of the proof contains the estimate for $[\varepsilon^{\bar{s},\bar{\mathbf{u}}}_{\bar{i},\theta}\det Z_{\bar{i}}]^{-2n}$, and is yet another version of Lemma 5.2.

Lemma 6.1. *Let* Π *possess the wide cone condition and* $a\in\mathbf{K}_\infty$. *Let* $k\in\mathbb{N}$ *be fixed, and let the constant* B *in the construction of the grids* $\mathcal{G}^\gamma$ *be taken greater than* k. *Then, for every* $\alpha>0$ *under an arbitrary choice of the constant* A *in the construction of the grids* $\mathcal{G}^\gamma$,

$$\sup_\gamma\sup_{l\le k}\sup_{i_1,\dots,i_l\in\mathbb{N}}\sup_{\mathbf{u}_1\in\mathbf{\Gamma}_{i_1},\dots,\mathbf{u}_l\in\mathbf{\Gamma}_{i_l}}\sup_{s_1,\dots,s_l\in[0,t]}\left[\mathsf{E}^0_{i_1,\dots,i_l}[\varepsilon^{s_1,\mathbf{u}_1}_{i_1}\dots\varepsilon^{s_l,\mathbf{u}_l}_{i_l}\det Z_{(i_1,\dots,i_l)}]^{-\alpha}\right]<+\infty.$$

Proof. Again we consider only the case $k=1$, the general case is completely analogous. Like in the proofs of Lemmae 5.2,5.5, we return from the "censored" probability space (Ω^0_i, M^0_i) to the initial one (Ω,P) and estimate $E[\varepsilon^{s,\mathbf{u}}\det Z_i]^{-\alpha}$, where

$$Z_i=\sum_{\tau_k\in\mathcal{D},\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i} g(\tau_k)\otimes g(\tau_k),$$

$$g(\tau_k)\equiv Jh(\tau_k-[\tau_k])A^{[\tau_k]}(\|p(\tau_k)\|^{-1}\wedge 1)\mathcal{E}^0_{\tau_k}\Big[a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big)\Big]$$

Let N be fixed. We denote by $\mathcal{D}(i,N)$ the set of $\tau_k\in\mathcal{D}$ such that $\mathbf{p}(\tau_k)\notin\mathbf{\Gamma}_i$ and $\tau_k\in\bigcup_{r=1}^N[-r-\frac{1}{3},-r-\frac{2}{3}]$. Let us estimate the variable

$$\det Z_{i,N},\quad Z_{i,N}\equiv\sum_{\tau_k\in\mathcal{D}(i,N)} g(\tau_k)\otimes g(\tau_k).$$

It is clear that $\varepsilon^{s,\mathbf{u}}\det Z_{i,N}\le\varepsilon^{s,\mathbf{u}}\det Z_i$, thus the lower estimate for $\varepsilon^{s,\mathbf{u}}\det Z_{i,N}$ provides also the lower estimate for $\varepsilon^{s,\mathbf{u}}\det Z_i$.

We write the decomposition $g(\tau_k)=\mathcal{E}^0_{-N}q(\tau_k)$,

$$q(\tau_k)\equiv Jh(\tau_k-[\tau_k])A^{[\tau_k]}(\|p(\tau_k)\|^{-1}\wedge 1)[\mathcal{E}^{\tau_k}_{-N}]^{-1}\Big[a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big)\Big],$$

and define

$$Q_{i,N} = \sum_{\tau_k \in \mathcal{D}(i,N)} q(\tau_k) \otimes q(\tau_k).$$

Then $Z_{i,N} = \mathcal{E}^0_{-N} \cdot Q_{i,N} \cdot [\mathcal{E}^0_{-N}]^*$ and $\det Z_{i,N} = \det Q_{i,N} \cdot \Big[\det \mathcal{E}^0_{-N}\Big]^2$. It would be convenient for us to formulate all the estimates concerned to $\mathcal{E}^{\cdot}_{-N}$ in a separate statement.

Proposition 6.2. *The following estimates hold true almost surely for every $T > 0$:*
1) $\det \mathcal{E}^0_{-T} \geq \exp\{-mTC(a)\}$;
2) $\inf_{\|v\|=1} \|(\mathcal{E}^s_{-T})^*]^{-1} v\| \geq \exp\{-TC(a)\}$, $s \in [-T, 0]$;
3)$\|\mathcal{E}^0_{-T}\|, \|[\mathcal{E}^s_{-T}]^{-1}\| \in [\exp\{-TC(a)\}, \exp\{TC(a)\}]$.

Proof. The first estimate is implied by the representation

$$\det \mathcal{E}^0_{-T} = \exp\{\int_{-T}^0 \mathrm{trace}(\nabla a(X(s)))\, ds\}. \tag{6.7}$$

This representation follows from the same one for ODE's, that is a classical fact in theory of ODE's. In order to deduce (6.7) in the framework of the equations with the Lévy noise one should first prove (6.7) for a compound Poisson process U by just applying (6.7) for ODE's piecewisely and then use an approximation procedure.

In order to deduce the second estimate, we use the equality

$$[(\mathcal{E}^s_{-T})^*]^{-1} v = v - \int_{-T}^s [\nabla a(X(r))]^* [(\mathcal{E}^r_{-T})^*]^{-1} v\, dr$$

that implies that, for every $v \in S_m$, the function $V(s) = \|[(\mathcal{E}^s_{-T})^*]^{-1} v\|$ satisfies the inequality

$$V(s) \geq v - \int_{-T}^s C(a) V(r)\, dr, \quad s \geq -T. \tag{6.8}$$

Inequality (6.8) can be written in the form of the equation

$$V(s) = 1 + \Delta(s) - \int_{-T}^s C(a) V(r)\, dr$$

with the condition $\Delta(s) \geq 0$, and the solution to this equation can be given in the form

$$V(s) = \exp\{-(s+T)C(a)\} + \int_{-T}^s \exp\{-(s-r)C(a)\}\Delta(r)\, dr \geq \exp\{-(s+T)C(a)\}.$$

The last estimate follows from the Gronwall lemma, on the one hand, and from the arguments given in the proof of the second estimate, on the other hand. The proposition is proved.

One can see that the same estimates with those given made in Proposition 6.2 hold true for $\varepsilon^{s,\mathbf{u}}\mathcal{E}^{\cdot}_{-T}$. Due to statement 1), $\varepsilon^{s,\mathbf{u}} \det Z_{i,N} \geq \det \varepsilon^{s,\mathbf{u}} Q_{i,N} \cdot \exp\{-2mNC(a)\}$. Let us estimate $\varepsilon^{s,\mathbf{u}} \det Q_{i,N}$. In order to do this, we will appropriately modify the arguments given in the proof of Lemma 1 [18].

Due to the condition on Π, there exists $\varrho \in (0,1)$ such that $\Pi(V(w,\varrho)) = +\infty$ for every cone $V(w,\varrho), w \in S_m$. Let $a \in \mathbf{K}_r, r \in \mathbb{N}$, further we denote $D = [D(a,r,\varrho) \wedge 1]$. For every given $\Lambda > 0$, there exists $\delta = \delta(\Lambda, \varrho)$ such that

$$\Pi\Big(u | u \in V(w,\varrho), \|u\| < D, |(u,w)_{\mathbb{R}^m}| > \delta\Big) \geq \Lambda$$

(one can prove this using the Dini theorem analogously to Lemma 4.3).

We take an arbitrary $v \in S_m$, and denote by $\mathcal{D}(i,N,v,\delta)$ the subset of $\mathcal{D}$ containing all the points τ_k such that $\tau_k \in \bigcup_{r=1}^N [-r - \frac{1}{3}, -r - \frac{2}{3}]$, $\mathbf{p}(\tau_k)$ does not belong to the cell $\mathbf{\Gamma}_i$, and

$$\varepsilon^{s,\mathbf{u}} \left| \Big(\Big[a\Big(X(\tau_k-) + p(\tau_k)\Big) - a\Big(X(\tau_k-)\Big)\Big], [(\mathcal{E}^{\tau_k}_{-N})^*]^{-1} v\Big)_{\mathbb{R}^m} \right| \geq D\delta^r \cdot \varepsilon^{s,\mathbf{u}} \left\| [(\mathcal{E}^{\tau_k}_{-N})^*]^{-1} v \right\|_{\mathbb{R}^m}.$$

The same arguments with those used in the proofs of Theorem 1.1 and Lemmae 5.2, 5.5 provide that

$$P(\mathcal{D}(i,N,v,\delta)=\emptyset)\le \exp\{-\frac{B-1}{3B}\Lambda N\}.$$

For every $v,\tilde v\in S_m$, due to statement 3) of Proposition 6.2 we get for every $\tau_k\in\mathcal{D}(i,N,v,\delta)$

$$\varepsilon^{s,\mathbf{u}}\left|\left(a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big),[(\mathcal{E}^{\tau_k}_{-N})^*]^{-1}v\right)_{\mathbb{R}^m}-\right.$$

$$\left.-\left(a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big),[(\mathcal{E}^{\tau_k}_{-N})^*]^{-1}\tilde v\right)_{\mathbb{R}^m}\right|\le 2C(a)\|[(\mathcal{E}^{\tau_k}_{-N})^*]^{-1}\|\|v-\tilde v\|\le 2C(a)e^{C(a)N}\|v-\tilde v\|.$$

Let us choose the vectors $v_1,\dots,v_{\exp[2C(a)N]}$ on the sphere S_m in such a way that, for every $v\in S_m$, $\inf_{l\le\exp[2C(a)N]}\|v-v_l\|\le\exp[-(2C(a)+\frac12)N]$ (one can do this for N large enough). Consider the event

$$\Omega(i,N,\delta)=\bigcap_{l=1}^{\exp[2C(a)N]}\{\mathcal{D}(i,N,v_l,\delta)\ne\emptyset\},\quad P(\Omega(i,N,\delta))\ge 1-\exp\left\{\Big[2C(a)-\frac{B-1}{3B}\Lambda\Big]N\right\}.$$

Take $v\in S_m$ and $l\le\exp[2C(a)N]$ such that $\|v-v_l\|\le\exp[-(2C(a)+\frac12)N]$. Then, for every $\omega\in\Omega(i,N,\delta)$, there exists $\tau_k\in\mathcal{D}(i,N,v_l,\delta)\subset\mathcal{D}(i,N)$. For such τ_k, we have

$$\varepsilon^{s,\mathbf{u}}\left|\left(a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big),[(\mathcal{E}^{\tau_k}_{-N})^*]^{-1}v\right)_{\mathbb{R}^m}\right|\ge$$

$$\ge\varepsilon^{s,\mathbf{u}}\left|\left(a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big),[(\mathcal{E}^{\tau_k}_{-N})^*]^{-1}v_l\right)_{\mathbb{R}^m}\right|-2C(a)\exp[-(C(a)+\frac12)N]\ge$$

$$\ge D\delta^r\cdot\varepsilon^{s,\mathbf{u}}\left\|[(\mathcal{E}^{\tau_k}_{-N})^*]^{-1}v_l\right\|_{\mathbb{R}^m}-2C(a)\exp[-(C(a)+\frac12)N]\ge\left(D\delta^r-2C(a)\exp\Big[-\frac N2\Big]\right)\exp[-C(a)N]\tag{6.9}$$

(the last inequality in (6.9) holds true due to statement 2) of Proposition 6.2). Take N large enough for $D\delta^r-2C(a)\exp\Big[-\frac N2\Big]\ge\frac12 D\delta^r$. Then, due to the construction of the grid $\mathcal{G}^\gamma$ and inequality (6.9) for every $v\in S_m$, we have the estimate

$$\varepsilon^{s,\mathbf{u}}(Q_{i,N}v,v)_{\mathbb{R}^m}\ge\sum_{\tau_k\in\mathcal{D}(i,N)}A^{-2N}\varepsilon^{s,\mathbf{u}}\left(a\Big(X(\tau_k-)+p(\tau_k)\Big)-a\Big(X(\tau_k-)\Big),[(\mathcal{E}^{\tau_k}_{-N})^*]^{-1}v\right)^2_{\mathbb{R}^m}\ge$$

$$\ge\frac14 D^2\delta^{2r}A^{-2N}\cdot e^{-2C(a)N}\mathbf{I}_{\Omega(i,N,\delta)}.$$

Thus, for every $\omega\in\Omega(i,N,\delta)$, we have the estimate

$$\varepsilon^{s,\mathbf{u}}\det Z_{i,N}\ge[\varepsilon^{s,\mathbf{u}}\inf_{v\in S_m}(Q_{i,N}v,v)]^{-m}e^{-2mC(a)N}\ge\frac{1}{4^m}D^{2m}\delta^{2rm}\exp[-C(a,A,m)N],\tag{6.10}$$

$C(a,A,m)\equiv 2m[\ln A+2C(a)]$.

At last, take Λ large enough for $\frac{B-1}{3B}\Lambda-2C(a)>(\alpha+1)C(a,A,m)$ and consider the sequence $t_N=\frac{1}{4^m}D^{2m}\delta^{2rm}\exp[-C(a,A,m)N],N\ge1$ (recall that δ is defined by Λ). Then (6.10) provides that, for N large enough,

$$P(\varepsilon^{s,\mathbf{u}}\det Z_i\le t_N)\le P(\varepsilon^{s,\mathbf{u}}\det Z_{i,N}\le t_N)\le 1-P(\Omega(i,N,\delta))\le C_\bullet\,[t_N]^{\alpha+1}.$$

Since $t_N\to0+$ with $\limsup_N\frac{t_N}{t_{N+1}}<+\infty$, this completes the proof of the lemma. The lemma is proved.

Corollary 6.1. *Let N be fixed. Then under conditions of Theorem 1.5 one can construct the grids $\mathcal{G}^\gamma$ in such a way that, for every $\boldsymbol{\alpha}\in\{1,\dots,m\}^n,n\le N$, the integration-by-parts formula (6.4) holds true with*

$$\sup_\gamma\sum_{\theta\in\Theta(2n)}\sum_{\bar\imath\in\mathbb{N}_d^{2n}}\int_{\mathbf{\Gamma}_{\bar\imath}}\int_{S(\bar\imath,\theta)}\mathsf{E}^0_{\bar\imath}|Y^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar\imath,\theta}(\bar s)|\,\lambda_{\bar\imath,\theta}(d\bar s)\,\mu_{\bar\imath}(d\bar{\mathbf{u}})=\mathsf{C}_n<+\infty.\tag{6.11}$$

Proof. In estimate (6.6), the term $\min(s_1, \dots, s_{2n})$ can be replaced by $\max(N(i_1, \dots, N(i_{2n}))$. Now, let us take the constant A in the construction of the grid to be equal to $2e^{C(a)}$. Then Lemma 6.1 and estimates (6.5),(6.6), together with the Hölder inequality, provide that

$$\int_{\mathbf{\Gamma}_{\bar{i}}} \int_{S(\bar{i},\theta)} \mathsf{E}^0_{\bar{i}} |Y^{\bar{\mathbf{u}},\boldsymbol{\alpha}}_{\bar{i},\theta}(\bar{s})| \, \lambda_{\bar{i},\theta}(d\bar{s}) \, \mu_{\bar{i}}(d\bar{\mathbf{u}}) \le$$

$$\le C_\bullet \, \lambda_{i_1} \dots \lambda_{i_{2n}} (\varepsilon_{n(i_1)} \wedge 1) \dots (\varepsilon_{n(i_{2n})} \wedge 1) C_\bullet \, (1 + \|\varepsilon_{n(i_1)}\|) \dots (1 + \|\varepsilon_{n(i_{2n})}\|) 2^{-\max(N(i_1), \dots, N(i_{2n}))},$$

$\bar{i} \in \mathbb{N}^{2n}_d, \theta \in \Theta(2n)$. Taking the sum over $\bar{i}, \theta$ we obtain (6.11).

End of the proof of Theorem 1.5. The corollary given above implies that, for every given $k \in \mathbb{N}$, one can construct the grids $\mathcal{G}^\gamma$ in such a way that estimates (5.7) hold true for every $n \le k + m$. Thus, due to Lemma 5.1, $P^*(dy) = p^*(y)dy$ with $p^* \in \bigcap_k CB^k(\mathbb{R}^m) = C^\infty_b(\mathbb{R}^m)$. The theorem is proved.

Acknowledgement. The technique exposed in Section 6 and the result of Theorem 1.5 were motivated by the question of A.A. Dorogovtsev, who asked the author whether the stochastic calculus of variations, described in the Section 3, can provide the smoothness of the invariant density. It was a quite new idea for the author that the differential properties of the invariant distribution can be essentially different from those of the distribution of the solution to the Cauchy problem. The author would like to express his gratitude to A.A. Dorogovtsev for the question that opened for the author a new and fruitful research field.

References

[1] *D.E. Alexandrova, V.I. Bogachev, A.Yu.Pilipenko.* On convergence in variation of the induced measures // Mat. Sbornik. – 1999. – v. 190, N 9. – P.3-20. (in Russian).

[2] *K.Bichteler, J.-B.Gravereaux, J.Jacod.* Malliavin calculus for processes with jumps. – New York, Gordon and Breach, 1987.

[3] *J.M.Bismut.* Calcul des variations stochastiques et processus de sauts // Zeit. fur Wahr. – 1983. – 63. – 147 – 235.

[4] *V.I.Bogachev.* Differentiable measures and the Malliavin calculus // J. of Math. Sci. – 1997. – Vol. 87, N 5. – P.3577-3731.

[5] *E.Carlen, E.Pardoux.* Differential calculus and integration-by-parts on Poisson space // Stoch. Algebra and Analysis in Classical and Quantum Dynamics (Marseille, 1988). – Math. Appl. – 1990. – vol. 59. – 63-67.

[6] *Yu.A.Davydov, M.A.Lifshits.* Stratification method in some probability problems. – Prob. Theory, Math. Statist., Theor. Cybernetics. – 1984. – vol 22. – 61-137. (in Russian).

[7] *Yu.A.Davydov, M.A.Lifshits, N.V.Smorodina.* Local Properties of Distributions of Stochastic Functionals. – Moscow, Nauka, 1995. (in Russian).

[8] *R.J.Elliott, A.H.Tsoi.* Integration-by-parts for Poisson processes // J. of Multivar. Analysis. – 1993. – Vol. 44 – P.179-190.

[9] *G.Federer.* Geometric Measure Theory. – Nauka, Moscow. – 1987. (Russian, translated from G.Federer, Geometric Measure Theory. – New York, Springer, 1969).

[10] *I.I.Gikhman, A.V.Skorokhod.* Stochastic Differential Equations and Their Applications. – Kiev, Naukova Dumka, 1982. (in Russian).

[11] *N.Ikeda, S.Watanabe.* Stochastic Differential Equations and Diffusion Processes. – Amsterdam, North-Holland, 1981.

[12] *Y.Ishikawa.* Density estimate in small time for jump processes with singular Lévy measures // Tohoku Math. J. – 2001. – Vol. 53. – P.183-202.

[13] *Y.Ishikawa, H.Kunita.* Existence of density for canonical differential equations with jumps// Preprint Nanzan. – 2004. O.

[14] *O.Kallenberg.* Splitting at backward times in regenerative sets// Ann. Probab. – 1981. – Vol. 9. – P. 781-799.

[15] *A.N.Kochubei.* Parabolic pseudodifferential equation, supersingular integrals and Markov processes // Izv. Akad. Nauk, Ser. Mat. – vol. 52, N.2. – 1982. – 909 – 934 (in Russian); Math.-USSR Izv. – vol 33. – 1983. – 233 – 259 (in English).

[16] *V.N.Kolokoltsov.* Symmetric stable laws and stable-like jump-diffusions // Proc. London Math. Soc. – 2000. – vol. 80, N 3. – P.725-768.

[17] *T.Komatsu, A.Takeuchi.* On the smoothness of PDF of solutions to SDE of jump type // Int. J. Differ. Equ. Appl. – 2001. – vol 2, N2. – 141–197.

[18] *T.Komatsu, A.Takeuchi.* Simplified probabilistic approach to the Hórmander theorem // Osaka J. Math. – vol. 38. – 2001. – 681 – 691.

[19] *A.M.Kulik.* Admissible transformations and Malliavin calculus for compound Poisson process // Theory of stochastic processes. – 1999. – Vol. 5(21), N 3-4. – 120-126.

[20] *A.M.Kulik.* Markov uniqueness and Rademacher theorem for smooth measures on infinite-dimensional space under successful filtration condition // Ukr. Math. J. – 2005. – vol. 57, N 2. – P.170-186.

[21] *A.M.Kulik.* Some remarks on time-stretching differentiation for general Lévy processes // Theory of Stochastic Processes. – 2001. – Vol. 7(23), N 3-4. – P.50-63.

[22] *A.M.Kulik.* Malliavin Calculus for Lévy Processes With Arbitrary Lévy Measures // Prob. Theor. Math. Stat. – 2005. – vol. 72. – 67-83.

[23] *A.M.Kulik.* On a regularity of distribution for solution of SDE of a jump type with arbitrary Lévy measure of the noise //Ukr. Math. J. – 2005. – vol. 57, N 9. – 1261-1283.

[24] *A.M.Kulik.* On a convergence in variation for distributions of solutions of SDE's with jumps // Random Operators and Stoch. Equations. – 2005. – Vol.13, N 3. – P.297-312.

[25] *R.Léandre.* Regularites de processus de sauts degeneres (II). – Ann. Inst. Henri Poincare Prob. Stat. –1988. – vol.24. – 209-236.

[26] *P.Malliavin.* C^k-hypoellipticty with degeneracy. – Stochastic Analysis. – New York, Acad. Press, 1978. - vol. 199-214. – 327-340.

[27] *J.Norris.* Simplified Malliavin Calculus // Lecture Notes in Math. – 1985. – vol. 1206. – 101-130.

[28] *I.Nourdin, T.Simon.* On the absolute continuity of Lévy processes with drift // Preprint. – 2005.

[29] *J.Picard.* On the existence of smooth densities for jump processes // Probab. Theory Rel. Fields. – 1996. – Vol. 105. – P.481-511.

[30] *A.Yu.Pilipenko.* Properties of stochastic differential operators in non-Gaussian case // Theory of Stochastic Processes. – 1996. – vol 2(18), N. 3-4. – 153-161.

[31] *N.Privault.* Linear Skorohod differential equations on Poisson space // Stochastic Analysis and Related Topics V, Progress in Probability. – Basel, Birkhauser, 1996. – Vol. 38. – P.237-253.

[32] *N.Privault.* Equivalence of gradients on configuration spaces // Random Oper. Stoch. Eq. – 1999. – Vol. 7, N 3. – P.241-262.

[33] *N.Privault.* Connections and curvature in the Riemannian geometry of configuration spaces // J. of Func. Anal. – 2001. – Vol. 183. – P.367-403.

[34] *P.E.Protter.* Stochastic Integration and Differential Equations. – Applications of Mathematics, Stochastic Modelling and Applied Probability. – Berlin, Springer, 2004. – Vol. 21.

[35] *H.Rubin.* Supports of convolutions of identical distributions // Proceedings of the 5th Berkeley Symposium on Math. Stat. and Probability, vol. 2, part 1, Univ. Calif. Press. – 1967. – P.415-422.

[36] *K.I.Sato.* Absolute continuity of multivariate distributions of class L // J. Multivariate Anal. – 1982. – Vol. 12. – P.89-94.

[37] *K.I.Sato.* Time evolution of distributions of Lévy processes from continuous singular to absolutely continuous // Research Bulletin, Ser. B, The College of General Education, Nagoya University. – No. 38. – 1994. – P.1-11.

[38] *A.V.Skorokhod.* Random Processes with Independent Increments. – Moscow, Nauka, 1967. (in Russian).

[39] *H.G.Tucker.* On a necessary and sufficient condition that an infinitely divisible distribution be absolutely continuous // Trans. Amer. Math. Soc. - 1965. – Vol. 118. – P.316-330.

[40] *S.Watanabe.* Lectures on Stochastic Differential Equations and Malliavin Calculus. — Tata Inst. Fund. Res. – Berlin, Springer, 1984.

[41] *M.Yamazato.* Absolute continuity of transition probabilities of multidimensional processes with independent increments // Probab. Theor. Appl. – 1994. – vol. 38, N.2. – P. 422-429.

Institute of Mathematics, Ukrainian National Academy of Sciences, 3, Tereshchenkivska Str., Kyiv 01601, Ukraine

E-mail address: `kulik@imath.kiev.ua`